Report on

High Performance Computing for the National Security Community

1 July 2002

OFFICE OF THE SECRETARY OF DEFENSE

1000 DEFENSE PENTAGON
WASHINGTON, DC 20301-1000

APR 9 2003

The Honorable Duncan L. Hunter
Chairman, Committee on Armed Services
United States House of Representatives
Washington, DC 20515-6035

Dear Mr. Chairman:

As requested within the House of Representatives Report 107-298, Department of Defense Appropriations Bill, 2002 and Supplemental Appropriations, 2002, page 211, the enclosed report titled "High Performance Computing for the National Security Community" is submitted. This report outlines a strategy currently under evaluation by the Department to rebuild and sustain a robust technology and industrial base in high-end supercomputing. This strategy builds upon current Department research and development efforts in this critical dual-use technology area. Over the next six months, the Department will develop an appropriate implementation plan for high-end computing, working in close coordination with other government agencies.

A similar letter has been sent to the House and Senate Appropriations Committees and to the Senate Armed Services Committee.

Sincerely,

E. C. Aldridge
Under Secretary of Defense for
Acquisition, Technology and
Logistics

John P. Stenbit
John P. Stenbit
Assistant Secretary of Defense for
Command, Control, Communications
and Intelligence

Enclosure:
As stated

cc:
The Honorable Ike Skelton
Ranking Member

OFFICE OF THE SECRETARY OF DEFENSE

1000 DEFENSE PENTAGON
WASHINGTON, DC 20301-1000

APR 9 2[illegible]

The Honorable Jerry Lewis
Chairman, Committee on Appropriations
Subcommittee on Defense
United States House of Representatives
U.S. Washington, DC 20515-6018

Dear Mr. Chairman:

As requested within the House of Representatives Report 107-298, Department of Defense Appropriations Bill, 2002 and Supplemental Appropriations, 2002, page 211, the enclosed report titled "High Performance Computing for the National Security Community" is submitted. This report outlines a strategy currently under evaluation by the Department to rebuild and sustain a robust technology and industrial base in high-end supercomputing. This strategy builds upon current Department research and development efforts in this critical dual-use technology area. Over the next six months, the Department will develop an appropriate implementation plan for high-end computing, working in close coordination with other government agencies.

A similar letter has been sent to the Senate Appropriations Committee and to the House and Senate Armed Services Committees.

Sincerely,

E. C. Aldridge
Under Secretary of Defense for
Acquisition, Technology and
Logistics

John P Stenbit

John P. Stenbit
Assistant Secretary of Defense for
Command, Control, Communications
and Intelligence

Enclosure:
As stated

cc:
The Honorable John P. Murtha
Ranking Member

OFFICE OF THE SECRETARY OF DEFENSE

1000 DEFENSE PENTAGON
WASHINGTON, DC 20301-1000

APR 9 2[illegible]

The Honorable Ted Stevens
Chairman, Committee on Appropriations
Subcommittee on Defense
United States Senate
Washington, DC 20510-6028

Dear Mr. Chairman:

As requested within the House of Representatives Report 107-298, Department of Defense Appropriations Bill, 2002 and Supplemental Appropriations, 2002, page 211, the enclosed report titled "High Performance Computing for the National Security Community" is submitted. This report outlines a strategy currently under evaluation by the Department to rebuild and sustain a robust technology and industrial base in high-end supercomputing. This strategy builds upon current Department research and development efforts in this critical dual-use technology area. Over the next six months, the Department will develop an appropriate implementation plan for high-end computing, working in close coordination with other government agencies.

A similar letter has been sent to the House Appropriations Committee and to the Senate and House Armed Services Committees.

Sincerely,

E. C. Aldridge
Under Secretary of Defense for
Acquisition, Technology and
Logistics

John P Stenbit

John P. Stenbit
Assistant Secretary of Defense for
Command, Control, Communications
and Intelligence

Enclosure:
As stated

cc:
The Honorable Daniel K. Inouye
Ranking Member

OFFICE OF THE SECRETARY OF DEFENSE

1000 DEFENSE PENTAGON
WASHINGTON, DC 20301-1000

APR 9 2003

The Honorable John Warner
Chairman, Committee on Armed Services
United States Senate
Washington, DC 20510-6050

Dear Mr. Chairman:

As requested within the House of Representatives Report 107-298, Department of Defense Appropriations Bill, 2002 and Supplemental Appropriations, 2002, page 211, the enclosed report titled "High Performance Computing for the National Security Community" is submitted. This report outlines a strategy currently under evaluation by the Department to rebuild and sustain a robust technology and industrial base in high-end supercomputing. This strategy builds upon current Department research and development efforts in this critical dual-use technology area. Over the next six months, the Department will develop an appropriate implementation plan for high-end computing, working in close coordination with other government agencies.

A similar letter has been sent to the Senate and House Appropriations Committees and to the House Armed Services Committee.

Sincerely,

E. C. Aldridge
Under Secretary of Defense for
Acquisition, Technology and
Logistics

John P Stenbit

John P. Stenbit
Assistant Secretary of Defense for
Command, Control, Communications
and Intelligence

Enclosure:
As stated

cc:
The Honorable Carl Levin
Ranking Member

EXECUTIVE SUMMARY

Introduction

The Report accompanying the House FY02 Defense Appropriations Bill directed the Secretary of Defense to submit a development and acquisition plan for a comprehensive, long-range, integrated, high-end computing program to Congress by July 1, 2002. This report presents that plan. It was prepared by the National Security Agency (NSA) in cooperation with the Defense Advanced Research Projects Agency (DARPA), the Department of Defense's (DOD) High Performance Computing (HPC) Modernization Program, the National Reconnaissance Office (NRO), the National Nuclear Security Administration (NNSA) of the Department of Energy (DOE), and the National Aeronautics and Space Administration (NASA). The plan incorporates the best ideas from the experts in High-End Computing (HEC) at the participating agencies as well as their counterparts in academia, industry, national laboratories, and other parts of the government.

In preparing this report the agencies formed three working groups: the Operational User Working Group (OUWG) addressed and catalogued the high-end computing needs of priority national security programs; the Systems, Architecture, Programmability, and Components Working Group (SAPCWG) assessed the state of current and future high-end computing and, the Planning Working Group (PWG) created this report. A three member executive panel oversaw all study group activities. It included the Deputy Undersecretary of Defense for Science and Technology; the Director of Advanced Simulation and Computing Office at the National Nuclear Security Administration; and the Information Assurance Director of the National Security Agency who served as panel chair. A list of participants is provided in Appendix A.

Rationale

Throughout the entire period of the cold war, and the decade that followed, high performance computing made major contributions to the nation's security. It has been a key technology in the development of our nuclear arsenal, other major weapons systems, in critical intelligence fields of image processing and cryptanalysis, and as a key enabler for U.S. leadership in national security-related fields of science and discovery. National security interests largely created the supercomputing industry and until the mid-90's, drove it technologically. This has now changed. Non-defense industrial, scientific and academic markets for systems have grown significantly and, coupled with emergence of foreign competition for market share as well as technology leadership, have diluted the national security community's dominance. Given these trends over the past decade, both government and industry have focused on developing and manufacturing high-end supercomputers based upon commodity components. While this approach has significantly increased the affordability for solving many important national security computational problems, there still exist critical applications for the community that are neither met nor addressed by the commercial sector.

Previous studies sponsored by the Defense Department have confirmed the need for investments in high-end supercomputing focused at addressing important national security problems. A mid–90's IPT[1] concluded that a major national security high-end computing program was necessary for the US to maintain supremacy in the field; challenged at that point in time by aggressive Japanese marketing

Continued

[1] "DDRE Integrated Process Team Study - A National Security High End Computing Program" 1996.

campaigns. A Defense Science Board study[2] supported the ongoing joint DDRE/NSA effort to develop the CRAY SV2, now being introduced to the market. This study and a MITRE study[3] provided the supporting rationale for the DARPA High Productivity Computing (HPCS) Program. However, these initiatives are one-time efforts aimed at developing the next generation of systems (i.e., Cray SV-2 and the HPCS in 2008 to 2010) and will not sustain the steady stream of new technologies and concepts necessary for the continued advancement of high-end computing. In addition, the basic research portfolio of the National Security community is too small and fragmented to maintain essential support to a comprehensive program. High Performance Computing R&D across the government is also skewed too heavily towards research related to quantum computing, at the expense of other promising and more realizable computing technologies[4]. Furthermore, the DOE ASCI Path Forward Initiative, which invests in accelerating industry technology efforts (e.g., switches) has seen significant reductions in budget.

For the working groups involved with this report, the situation is clear. The mix of research, development, and engineering programs lack balance and coordination and is far below the critical mass required to sustain a robust technology/industrial base in high-end supercomputing. Requirements identified as critical by the national security user community (such as improved memory subsystem performance and more productive programming environments) will not be addressed. The impact is that the national security community will be unable to solve critical computational problems required to maintain our technology lead for select but important classes of problem, examined in the course of the study, which included:

- Weapons Development Program
  - Comprehensive Air Vehicle Design
  - Army Future Combat Systems
  - Stealthy Ship Design
- Nuclear Stockpile Stewardship
- Cryptanalysis
- Global Ocean Modeling and Operational Fleet Weather Forecasting
- Biological Sciences
- Intelligence Support
  - Threat Systems M&S
  - Signals & Image Proc
  - Nuclear Effects
- Future Critical Problems
- Missile Defense

An important example of a loss of technology leadership is the Japanese Earth Simulator introduced in March 2002. The capabilities of this machine need further verification, however, the machine offers the Japanese climate modeling community a capability orders of magnitude greater than those available to their U.S. counterparts. This system has little direct impact on U.S. national security interests, however, as a scalable, general purpose high end computer, it represents a potential threat to U.S. HEC companies and a

Continued

2 "Task Force on DoD Supercomputing Needs", Defense Science Board Study, October 11, 2000

3 "Survey and Analysis of the National Security High Performance Computing Architectural Requirements" Presentation by Dr. Richard Games, MITRE, April 26, 2001

4 Additional funding could be redirected from the $60M being spent on research related to quantum-computing. The panel of experts found this level of funding to be out of balance relative to the other HEC activities.

challenge to U.S. leadership in important scientific computational domains. The Earth Simulator demonstrates that other nations are capable of building systems of greater capacity and capability for specific problems of national interest. This system, built via a long-term GOJ/NEC program (though largely funded by the government), is an important example of what can be achieved through a sustained long-term cooperative investment program in this field.

Program Outline

To *rebuild and sustain* a strong industrial base in high-end supercomputing, the working groups recommended a coordinated spiral model program comprising the following elements: applied research, advanced development, and engineering and prototype development. The *applied research* element will focus on developing the fundamental concepts in high-end computing and creating a pipeline of new ideas and graduate-level expertise for employment in industry and the national security community. The *advanced development* element will select and refine innovative technologies and architectures for potential integration into high-end systems. The *engineering and prototype development* will build operational prototypes (i.e., "serial number 1") and system level testbeds. We foresee academic and industry participating and partnering in all elements, with universities serving as the focus of applied research and industry increasingly driving the latter two elements.

Furthermore, *high-end computing laboratories are needed*. These laboratories will fill a critical capability gap identified by the workings groups to test system software on dedicated large-scale platforms, support the development of software tools and algorithms, develop and advance benchmarking and modeling, and simulations for system architectures, and conduct detailed technical requirements analysis. It is envisioned that these functions would be executed by existing computer centers (both government and academic centers would be considered).

Our overarching strategy is to underwrite only that research, development and engineering that industry will not conduct. Our end-game is to leverage the U.S. computing industry to be market-sustainable but with minimum risk to long-term high end computing interests of the national security sector.

These three elements and the laboratories would be executed as integrated program, called the **Integrated High-End Computing (IHEC) Program**. This program will be constructed from a consolidation of existing DARPA, DOE/NNSA, and NSA programs and will require additional funding to provide for the research, development, and engineering needed to build future high-end computing systems. As described in this report, two options for additional funding are provided: a base option growing to require $250M per year and a progressive-level program eventually requiring $390M per year.

The IHEC program will be executed by a Joint Program Office and staffed by the participating national security agencies. Program execution would be through the R&D agencies (i.e., DARPA, DOE/NNSA, NASA, and NSA). Program oversight would be provided from the Office of the Director, Defense Research and Engineering. A senior level advisory group consisting of agency executives would serve to balance organizational needs and investment strategies. The implementation of this program would be finalized over the next year. Issues of transitioning existing programs at appropriate milestones and consolidating budgets will involve considerable interagency coordination.

Continued

EXECUTIVE SUMMARY, Con't.

It is important to note that agency high-end supercomputing procurements were not included as an element within this program. Within many of the agencies participating in this report, procurements are highly classified and focused at mission requirements and systems. The consensus of the working groups was that while coordination of research, development, and engineering investments was tractable and reasonable, procurements were not[5]. Therefore, an important premise of this program is that through the active participation of the national security community in all phases of the IHEC Program, a pipeline of systems will be created that meet Agency requirements. And consequently, the participating Agencies will procure these systems to meet their specific needs.

Conclusion

This report details a multi-agency program in high-end supercomputing in response to FY 2002 House Defense Appropriation Bill[6]. The proposed program will conduct a range of activities, from applied research to support for building prototype systems. The end-state of this program is a robust and viable industry base for high-end computing enabling the national security community to solve critical large-scale computational problems-both today and for the foreseeable future.

---

[5] Undercapitalization of operational high end computing capabilities was raised as a concern by a number of national security community user organizations participating in this study; these issues will need to be addressed separately.

[6] (Extract from the Report to Accompany FY02 Defense Appropriation Bill, p. 211. "Next Generation Supercomputing Capability.")

"The committee is concerned that several critical national security technologies suffer from inadequate Department of Defense (DoD) high-end supercomputing resources. These technologies include cryptanalysis, operational weather forecasting, dispersion of airborne contaminants, armor design of large aircraft or ship structures and studies of weapons effects.

"Over the last decade, declining markets, inequitable trade practices and limited DoD support have severely weakened the United States industrial base for high-end supercomputing. Several reports conducted at the request of the DoD have clearly identified a number of issues that counsel immediate attention and action to avoid an unacceptable prospect; offshore reliance for critical supercomputing capability vital to our national security.

"The committee directs the Security of Defense to submit by July 1, 2002, a development and acquisition plan, including budgetary requirements, for a comprehensive, long-range, integrated high-end supercomputing program. The secretary shall direct the National Security Agency (NSA) to take lead in developing this plan in cooperation with DARPA, DoD's HPC Modernization Program, NIMA, NRO, NNSA/ASCI, and NASA."

Contents

What is a High-End Supercomputer?

For the purposes of this study, high-end supercomputers (HEC) are defined as general-purpose computers with leading edge performance for some problem set. They achieve this performance by carefully balancing some combination of the following properties:

* Many interconnected processing elements
* Large memory
* Fast access memory
* High system bandwidth
* Low system latency
* Large capacity input/output
* High computational capability
* Massively parallel system software and programming environment

In their assessment of the current state of HEC, the SAPCWG found there to be two distinct classes of HEC systems. The first category, type-T HEC systems[7], is very efficient working applications with a high ratio of computation to communication. These HEC systems use the same highly refined and mass-produced processors as the $80B server and workstation market. Thus, the capabilities of type-T systems are expected to continue to improve, as their performance tends to track or exceed the phenomenal rate of evolutionary improvement the world has come to expect in mass-produced processors.

The other category, type–C systems, feature computer systems designed and developed to provide high communication bandwidth to global memory, and between processors. Type–C systems can offer significantly better performance for those classes of HEC applications with significant rates of random memory access and low ratios of computation to communications. However, this category of systems typically optimizes global communications by sacrificing some performance for those applications that can leverage data caches. This makes type–C systems less attractive for many server-based applications and consequently, they are manufactured in significantly smaller volumes.

National Security Needs for High–End Computing

The OUWG studied the HEC needs of system users in a variety of critical national security technologies and compared them to current and anticipated HEC capabilities.[8] It found that many of the advantages the U.S. enjoys in technologies critical to national security depend to a substantial degree on the relative strength and diversity of its domestic commercial sources for high-end computing.

Specifically, the OUWG found the U.S. uses its most powerful HEC systems today to achieve improved capabilities in a number of technologies—including many of those critical to national security. These uses are both necessary and prudent. The group identified the following list, in no particular order, of national security technologies that rely on HEC for both current and future needs. And, only time constraints kept the OUWG from identifying many more requirements.

Comprehensive Aerospace Vehicle Design—for simulation and modeling of advanced

[7] Burton Smith talk at CAS2001 in France

[8] "Report on High Performance Computing Requirements for the National Security Community," Operational Requirements Working Group, June 19, 2002.

fighters, cruise missiles, and reconnaissance aircraft extending into hypersonic velocities.

Signals Intelligence—for transformation, cryptanalysis, and intelligence analysis of foreign communications on the intentions and actions of foreign governments, militaries, espionage, sabotage, assassinations, or international terrorism.

Operational Weather & Ocean Forecasting—to provide worldwide 24-hour weather guidance to the military, CIA, and Presidential Support Unit for current operations, weapons of mass destruction contingency planning, etc.

Stealthy Ship Design—to allow future naval ships to accommodate emerging technologies such as electric propulsion systems and stealth to reduce detection signatures.

Nuclear Weapons Stockpile Stewardship—for the development and validation of high quality physics simulations of nuclear weapons to be used for stockpile certification in the absence of nuclear testing.

Signal and Image Processing—to find high value targets through the analysis of multispectral signals and images and the processing of more and improved sensor data at higher resolution in less time.

Army Future Combat Systems—to provide for rapid response and deployment of forces using advanced logistical analysis, as well as the discovery, refinement, development, and evaluation of critical lethality and survivability technologies.

Electromagnetic Weapons Development—for the development of the airborne laser for ballistic missile defense and other electromagnetic devices to disable battlefield electronics and provide alternatives to lethal weaponry.

Geospatial Intelligence—for the analysis and visual representation of security-related activities in reference to the Earth.

Threat Weapon Systems Characterization—for the development and application of advanced modeling, computational analysis, and simulation methodologies to the study and characterization of foreign threat weapon systems, as well as for the prediction of the performance of those threat systems in complex operating environments.

The OUWG found a significant subset of the national security community's current needs is being met. Nonetheless, as is evident from this partial listing, the HEC needs of the national security community are large and varied. And to meet the diversity of these needs, both type-T and type-C HEC systems are, and will continue to be, required by the national security community.

However, satisfying this diversity of requirements may prove problematic as today's server dominated market is forcing the major vendors to make design tradeoffs that do not address the full scope of the national security community's current and future HEC needs. This situation is resulting in a gap between HEC needs and capabilities. And, the trend identified by the team shows this gap widening over time.

The small number and size of the current initiatives ongoing in applied research, advanced development and engineering in HEC is particularly troublesome. That is because innovation is born from these efforts. The OUWG found a number of computational characteristics to be desperately in need of innovation as they are already forming bottlenecks to the realization of needed capabilities and productivity. In particular, the OUWG cited a critical need for better memory subsystem performance: latency, bandwidth, and size. The team also identified needs for improved programming environments, CPU performance, and input/output subsystems.

Finally, the OUWG noted that HEC, including advances in hardware, software, and programmability, has revolutionized the manner in which the national security community addresses a wide variety of critical national challenges. Many national security programs that give the United States dominance over its adversaries could not exist in their present state without a strong domestic commercial source for High-End Computing. Indeed, for the foreseeable future, our leadership in these technologies – clearly critical to national defense and homeland security – will depend in part on our researchers and engineers having uninterrupted access to leading edge supercomputing capabilities.

The Role of High-End Computers in Scientific Advancement

In the most fundamental sense, science is simply the aspiration for increased understanding. This applies to the study of both physical phenomena and mathematical relationships. Today scientific understanding is advancing at an unprecedented rate. Thus, investigators at the cutting edge find themselves pushed relentlessly to evaluate theories with ever-finer resolution. And while it is true that theories are validated ultimately by physical observation or measurement, this has historically taken decades or longer to accomplish. In the interim, investigators depend on mathematical modeling and simulation to evaluate, refine, and advance their theoretical understanding. Because high-end mathematical modeling and simulation of large and complex systems is computationally intensive, high-end computers are a critical tool for leading investigators. Thus it can be argued that when made available to expert mathematicians, scientists, and engineers, high-end supercomputers can provide them with a higher level of scientific understanding than could be achieved otherwise. This increased insight can result in improved capabilities that could be either anticipated or unforeseen. The latter comes from the fact that new, advanced modeling and simulation capabilities can give knowledge that literally expands the realm of thought. Such insight, in turn, spawns fresh concepts and ideas. Either way, increased scientific understanding often translates into significant strategic advantages.

More to the point, having had access to HEC systems has given the U.S. a decisive edge in innovation. As was stated previously, the fact that U.S. engineers, scientists, and mathematicians have had access to the leading edge HEC has enabled us to retain dominance in a number of technologies vital to our national security. The national security community has translated this edge in innovation into strategic advantages for the U.S. Furthermore, if we lose this edge, we will eventually find ourselves existing on a technical level closer to that of our adversaries.

Japan's Earth Simulator and the World Marketplace

In April 2002, the Japanese switched on their Earth Simulator supercomputer and the system immediately claimed the title of "world's fastest computer." The Earth Simulator reached a performance level of 35 TFlops (trillion floating point operations per second) on the Linpack benchmark test—a well–known measure of floating point computation performance. Previously Lawrence Livermore Laboratory's ASCI White system, built by IBM, held the title with a Linpack performance of 7 TFlops. More impressively, the Earth Simulator has achieved 25 TFlops sustained performance on a full physics atmospheric model, demonstrating the capability and efficiency of this system.

A consortium of Japanese government agencies[9] paid on the order of a half billion dollars to have NEC build the system. Design work began in 1997 and it was delivered on schedule. The Earth Simulator consists of 5,120 vector processors, grouped 8 to a node, with 640 nodes. The 640 nodes are all interconnected with a high–speed single stage crossbar. The system components were optimized for the specific problems it sets out to solve—climatic and tectonic predictions. And, it is believed that the Earth Simulator

[9] A joint effort of the National Space Development Agency of Japan, the Japanese Atomic Energy Research Institute and the Japanese Marine Science and Technology Center

provides Japanese earth scientists with a tool that is an order of magnitude faster than any other in the world.

The Earth Simulator demonstrates the degree of success that can be achieved through a long–term, sustained partnership between government, academia, and industry. Certainly, without the support of the Japanese government, NEC would not have developed the Earth Simulator. But, by sustaining their partnership across the interlinked phases of research, development, and first–system acquisition, the Japanese government and NEC have produced a highly capable system. Furthermore, by building the Earth Simulator, NEC is now well positioned to market smaller scale versions of the system (the NEC SX-6), with the non-recurring development costs having been covered by the Earth Simulator project. The U.S. has the technical capability to build such a system, however the type of partnership that the Japanese government has with NEC has not taken place in the U.S. at such a grand scale (the partnership between Cray Inc. and the U.S. government for the SV-2 is smaller scale).

The 2001 market for technical capability systems (as defined by IDC as multimillion dollar systems that are purchased to solve the largest, most demanding problems) is approximately $1 billion. And, three companies control 65% of this market. IBM currently holds the lead with 30% market share with its sales of very large clusters of thin node SMPs (Symmetric Multiprocessors). HP is second with a 20% market share with large system sales of the AlphaServer SC systems. NEC is third with a 15% share with its primary revenue coming from the manufacturing of the Earth Simulator.

The NEC Earth Simulator has shown dramatically that HEC systems with customized processors and a custom interconnect fabric can be optimized to perform well on global access of data, poorly balanced workloads, and adaptive or irregular meshes. IDC's current market forecast for technical capability computing predicts that NEC will have a strong growth in revenue with sales of the SX-6. NEC will have a strong presence in Europe and Japan and since Cray has signed a distribution agreement to sell NEC SX-6 systems within North America, there may also be a renewed presence in North America. Cray is the only U.S. vendor that will have a type-C system (the SV-2) with customized processors and a custom interconnect fabric that has been optimized to perform well on global access of data. IDC predicts that the SV-2 will have strong initial sales in the US, but Cray has a weak position in the global market, having only 4% of the market in 2001. The other US vendors market type-T HEC systems, with their mass-produced processors linked via fast interconnect fabrics, that perform well on applications with localized data and well-balanced workloads.

In summary, the state of HEC in the U.S. is dramatically different for type-T and type–C systems. Where the U.S. dominates the type–T market and should continue to do so, its presence in the type-C HEC market is precarious. And, the success of the Japanese government's partnership with NEC in producing the Earth Simulator presents the U.S. type-C market with a formidable technical challenge. However, this same partnership showed that this important industry—and the critical capabilities it provides—could be revitalized. To do this domestically, the U.S. must similarly commit to a partnership with academia and industry, and embark immediately on a plan with sustained long–term investment by the government in a research, development and acquisition program for high-end computing to meet critical future national security requirements.

The State of U.S. High-End Supercomputing

The SAPCWG investigated high performance computing for the national security community.[10] This group found that the semiconductor industry should continue to deliver faster and better processors and memory components. That is because the server and workstation market, which is assessed at $80B, will drive advancements in these areas. However, a note of caution was raised as research and development is focused mainly on technologies for the next one or two generations (18 to 36 months) of systems, and there is little long-term research by the computer industry. Conversely, as the worldwide market for high-end computers is approximately $1B, the outlook for the other segments of the high-end computing field is not good. In particular, the SAPCWG expressed a gloomy outlook for the following areas of the supercomputing field:

Programmability—Parallel programming will continue to get more difficult as systems scale to larger sizes. And, virtually no work is being done to rectify the situation.

System software—Many supercomputers run separate operating system images on each node. Cluster services are then built on top of the distributed operating systems. This results in poor system performance.

Input/Output—Current communication interfaces for moving data on and off chips throttle performance. This is discouraging as Storage Area Networks (SAN), Network Attached Storage (NAS), and parallel file system technologies mature. Within the decade these file systems can be expected to be able to deliver data at speeds far exceeding the rate at which chips can accept it.

Components and Basic Technology—The SAPCWG is concerned that the level of research and development in basic technology may be inadequate to continue to sustain system innovation much beyond five years. The study indicated the low number of scientists graduating from universities might be insufficient to support the research and development needs of the high-end system market.

Considering the overwhelming size of the server and workstation market, the team believes that U.S. manufacturers will increasingly concentrate on producing type-T high-end computers. In fact, the U.S. Government may actually be inadvertently accelerating this process. The U.S. National Science Foundation (NSF) is investing $53M to create its Teragrid supercomputer. The Teragrid project will interconnect supercomputers (each based on commodity processors) at four geographically dispersed locations into one virtual machine consisting of 3,300 processors with a peak performance of 13.6 TFlops (trillion floating point operations per second) and 600 Terabytes of data storage. This is certainly a worthwhile undertaking, and it unquestionably will benefit those applications that can have their problem split up into parts that can be worked individually and in parallel. However, the Teragrid project will also reinforce the trend toward building *all* high-end computers out of mass-produced processors. Consequently, the SAPCWG feels that without a complementary government development and acquisition program, type-C HEC systems will disappear altogether. If this is allowed to happen, some of the computational needs of critical national security technologies will be left unsatisfied.

This same team of investigators found that although national security community funding for applied research in fundamental HEC concepts has remained relatively constant, it has become too tightly focused on quantum computing and short–term research. This situation limits the areas being explored for potential innovation. Specifically, over the past five years, the fraction of HEC research funding spent by the national

10 "Report on High Performance Computing for National Security Community," Systems, Architecture, Programmability, and Components Working Group, June 19, 2002.

security HEC community on applied research in quantum computing grew from just a little over zero to approximately $60M. This imbalance has resulted in very little remaining to spend on applied research in areas outside of quantum computing. This situation has been exacerbated by industry's applied research being tightly focused on pushing for density improvements in silicon-based logic and memory devices.

In wrapping up their assessment of the state of HEC, the SAPCWG noted that the computational needs of technologies critical to national security are likely to continue to grow at a faster rate than is the development of new computational capabilities. And, given the current focus of the HEC industry, without government intervention this trend is expected to accelerate.

However, the SAPCWG also found that this unwanted trend can be reversed. The creative talent and skills needed to revitalize the high-end supercomputer industry are still resident in the USA. New, more powerful, and needed, capabilities can be developed with prompt action. Hence, the SAPCWG also made recommendations as to how the government should engage academia and industry to deliver significantly increased HEC capabilities. The team proposed a spiral model starting with applied technology research, proceeding through advanced development and prototyping phases leading to vendor products that support national security requirements.

Specifically, the SAPCWG recommended restoring the level and range of effort for applied research in fundamental HEC concepts to at least its original level, and with its original intent. This investment should be applied nearly evenly across seven general research areas: systems architectures; memory subsystems; parallel languages and programmer tools; packaging/power/thermal management; interconnects and switches; storage and input/output; and novel computational technologies (exclusive of quantum computing). The SAPCWG asserts that quantum computers will not displace conventional high performance computers. That is because, if they prove feasible, quantum computers are expected to be capable of producing exponential speedup for only a limited number of problem areas. Hence, quantum computing research should be treated as a separate activity outside the scope of this HEC program.

The SAPCWG also recommended increasing the number and level of advanced development efforts. This would facilitate the transition and combination of promising applied research ideas from the laboratory into subsystem prototypes and concept test beds and promote healthy competition amongst good ideas. Additionally, the team recommended increasing the number of engineering and operational prototype efforts supported by the national security community from one to at least two. The SAPCWG reasoned the broad diversity of the community's HEC needs could not be adequately served by a single development.

The team went on to suggest Time-To-Solution as an important concept in considering development and engineering efforts. For impressive as the Earth Simulator is, describing any supercomputer as the "world's most powerful system" is of questionable value. That is because the true measure of "power" for every user is relative. The system that yields a solution in the shortest amount of elapsed time for a particular application is the most powerful. And, for the most part in scientific computing, time-to-solution (TTS) should refer to the total amount of time required to write and debug the application program, compile it, as well as to execute it. TTS is influenced by system architecture, programming environment, component characteristics, and appropriateness of the system to the application.

Finally, it is important to place in context the proposed IHEC program in light of recent DoD high-end computing activities. Over the last several years, as reported in a number of DoD studies, there is a national security requirement for high productivity computing systems.[11,12] Findings were that without government R&D and participation, high end computing will be available only through commodity manufacturers primarily focused on mass-market consumer and business needs. This solution was found to be both ineffective for important national security applications and widely available to our potential adversaries. To fill the short-term DoD national security requirements, a development was initiated led by NSA to develop the next generation type-C high-end computer. Cray is now nearing completion of the SV2. Also as a result of these studies, a white paper in June 2001 was generated at the request of Principal Deputy Under Secretary of Defense for Acquisition, Technology, and Logistics to provide the DoD with increased long-term options for high end computing. A high-end computing program was defined with two primary goals. The first is to provide economically viable high productivity computing systems for the national security and industrial user community. The second is to reinvigorate the high-end hardware and software communities to develop a new generation of researchers, engineers, and leaders to drive the advancement and development of new high-end architectures and tools throughout the decade. In parallel with these efforts DARPA was exploring potential revolutionary concepts for longer-term high-end computing solutions. As result of all of these efforts, a major new program in high-end computing, High Productivity Computing Systems (HPCS) has been initiated by DARPA. Conceptual study awards have been made to CRAY, HP, IBM, SGI, and Sun. At the completion of this three-phase program, late in this decade, pre-production "Serial Number 1" high-end computing solutions are planned.

11 "Task Force on DoD Supercomputing Needs," Defense Science Board Study, October 11, 2000.

12 "Survey and Analysis of the National Security High Performance Computing Architectural Requirements," Presentation by Dr. Richard Games, MITRE, April 26, 2001.

What Needs to be Done

This study recommends that the U.S. government embark immediately on an Integrated HEC program with a clear vision and objectives to ensure a robust and sustained technology and industrial base for high-end computing. This integrated program, incorporating multi-agencies, would expand on the vision and current high-end programs already in place at DARPA, NNSA and NSA.

IHEC Vision: Provide improved national security capabilities by dramatically improving high-end computing.

IHEC Objectives:

Assure a healthy domestic high-end computing research and development environment and production capability focused on national security.

Develop the high-end computing technologies (hardware and software) necessary to advance national security initiatives that would not be produced by commercial interests.

Ensure commercial availability of high-end computing systems and technologies capable of providing the required time-to-solution for solving critical national security problems.

The IHEC program will involve a partnership between academia, industry, national laboratories and the national security mission agencies to conduct a wide range of activities—from applied research to building prototypes. The program will be structured across three technical elements: applied research, advanced development, and engineering and prototype development. Furthermore, a number of HEC laboratories would be supported to explore current HEC systems as well as to provide feedback for generating future capabilities. Figure 1 (next page) shows how these different elements work together and a detailed description of each element follows:

The *applied research* element will focus on the development of fundamental concepts in high-end computing. The intent is to create and maintain a continuous pipeline of new ideas and motivated graduate level expertise for employment in the computer industry and the national security community. Applied research activities are envisioned to be conducted primarily by academia with some industry participation. Projects will include small, long-term, high-risk efforts across a wide range of technologies and architecture concepts.

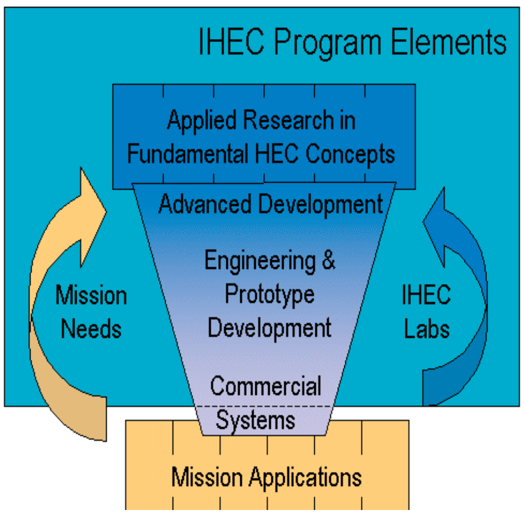


Figure 1-Conceptual view of the linkage between different IHEC Program elements.

The *advanced development* element will select and refine innovative technologies and architectures (some of which are developed under the applied research element) for potential integration into high-end systems. Subsystem prototypes and concept test beds will be built under this phase of the program. Projects will involve moderate-to-low component risk with a higher risk attributed to system integration. The intent of this element is to demonstrate critical hardware, software, and tool components in preparation for integration and prototype. Advanced development activities will be primarily led by industry with substantial academic participation.

The *engineering and prototype* element will provide operational prototypes and system level test beds. All of the engineering and testing costs will be provided to industry-led teams to produce "Serial Number 1" of a new system or component. The intent is that products via this phase of the program will be ready for either deployment or immediate commercialization.

HEC laboratories will provide several overarching functions for the program. These include the testing of system software on dedicated large-scale computer platforms; test and evaluation of prototype systems; support the development of tools and software algorithms for prototype systems; modeling and simulation to support concept and prototype development; and assist the program office in conducting detailed requirements analysis of national security applications. It is envisioned that one or more HEC laboratories will be established at existing computer centers (to include academic institutions).

It is significant to note that HEC system procurements are not, per se, part of this acquisition and development plan. That is because mission agencies must retain responsibility for procuring their HEC systems. Nonetheless, the PWG recognized that procurements ultimately provide the life-blood of the HEC industry as well as the fact that the national security community must acquire HEC systems to carry out its missions. Thus, instead of being an explicit part of the program, the IHEC will ensure that HEC systems produced by this program will meet the requirements of the mission agencies through their participation in program office functions, grant and contact proposal evaluations, and requirements analysis.

Applied Research in Fundamental High-End Computing Concepts and Technologies (ARHC)

A healthy high–end computing industry depends critically on a steady flow of new ideas from the fundamental areas of: systems architectures; memory subsystems; parallel languages and programmer tools; packaging/power/thermal management; interconnects and switches; storage and input/output; and novel computational technologies.

The Applied Research in Fundamental High-End Computing Concepts segment of the program produces the necessary ideas and concepts that will be used in subsequent higher-level phases of the program. This element will not only provide the mechanism responsible for producing the necessary ideas for future systems, but just as importantly, it will also produce the necessary U.S. citizen graduate students that will become the HEC scientists and engineers of tomorrow for the national security community.

The Program recognizes that while there is some component development going on in industry, it focuses on near-term (less than five years) timeframes. There is an insignificant level of research and development activity looking at greater than five years in the future. This lack of research and development, especially at universities, has produced a manpower gap; the United States is not producing the next generation of scientists that will conduct the research and development for basic technologies for the high-end computing industry going forward. In short, the research pipeline has drained and needs to be re-primed.

The principal outputs from this element of the program are ideas and techniques expressed in the form of technical papers, simulations, and hardware or software laboratory artifacts capable of demonstrating concepts.

In recognition of the importance of producing the next generation of scientists and engineers, this study recommends a scholarship program targeted at supporting U.S. citizen graduate students researching high-end computing. Limiting this graduate school scholarship program to U.S. citizens would eventually give the national security community a larger pool of HEC experts with which to support their sensitive operational needs.

The panel recommends a grants-and-contracts model of research funding to support basic high-end computing technologies focused at the long term and high risk. This research would be conducted primarily at U.S. universities, but also at industry and government research laboratories. A guiding principle of ARHC should be to deliver improvements in application time-to-solution.

It is anticipated that each area shall support several significant research projects on the order of one to two million dollars each, in addition to substantially more efforts ranging from $150,000 to $350,000 each to encourage early exploratory research.

Advanced Development (AD)

The planned high-end Advanced Development phase of the IHEC will logically grow out of DARPA's High Productivity Computing Systems (HPCS) initiative. As such, it will go through an initial start up phase, followed by a sustained stream of advanced development efforts. The initial advanced development phase was initiated as part of DARPA's HPCS program in 2001. The initial development consists of a concept study effort and a development phase. The one-year industry-led concept study initiated in 2002 will provide critical technology assessments, develop revolutionary HPCS concept solutions, and supply new productivity metrics necessary to develop a new class of high-end computers by the end of this decade. A major challenge is formulating a comprehensive set of formal requirements, benchmarking strategies, and metrics for these high-end tera- and peta-scale computing systems. Mitre Corporation is leading the applications analysis and performance assessment team. In 2003, the results from concept study will be merged with other DoD long-range high-end computing mission requirement assessments to form the basis for the first of a stream of four year advanced developments that will be initiated in late 2003. The advanced development phase will select and refine innovative technologies identified from the applied research efforts. The strategy is to first demonstrate critical HW/SW and tool components, modules, and architectures and then explore the vertical integration of critical HW/SW subsystem components. These pursuits will result in a series of system design reviews, preliminary design reviews and risk reduction prototypes and demonstrations. Developments of this type are best carried out by balanced industry, university, and research laboratory consortia. A goal is to replicate a number of these early subsystem prototypes to be placed in strategic universities and research centers for hands-on-research as a means to encourage and create a new generation of researchers. A down select of these efforts is planned for determining which of them will progress into the next phase: engineering and prototype development.

Engineering and Prototype Development (EPD)

The planned high-end EPD phase will go through an initial start up phase followed by sustained streams of EPD efforts. One of the models used to define the EPD phase was the NSA led, joint NSA, DARPA, and DoD effort which was begun in 1998, to acquire the best in class vector high performance computers for NSA's crucial national cryptanalysis mission. In that effort NSA and DoD entered into a long term Technology Investment Agreement with industry to perform the engineering and development necessary for bringing the Cray SV2 high-end computer to market in FY02. This approach was adopted only after it became clear that industry could not make a business case to develop a SV2 or equivalent HEC system entirely on its own. The decision to go forward was reviewed and validated by the Defense Science Board as a high payoff initiative for a modest expenditure.[13] The report further noted "supporting the SV2 might not be a one-time expense but rather a continuing investment in a critical defense-specific capability."[14] Consistent with this view, in FY02 NSA initiated a new effort with Cray, Inc. to stimulate and support the research and development of follow-on systems to the SV2.

The stream of sustained engineering and prototype development phases, to follow the initial efforts, will last four-to-five years and complete the detailed design, fabrication, integration and demonstration of the full-scale HEC pilots. The engineering and prototype development projects will be led by commercial industry. The goal is to provide early hardware and software releases or scaled-down replicas of the final pilot systems for early evaluation by academia and national security end users. Any actual pilot systems that are produced will be delivered and installed in HEC laboratories. Here the research community will be encouraged to evaluate HEC research platforms and software development environments, and to explore long-term, high-end computing challenges.

During the industry-led engineering and prototype development phase, full-scale pilot systems will be developed. The new HEC performance predictors and benchmarks developed earlier in the program will aid in long-term computing product procurements by providing an accurate measure of emerging commercial industry HEC products and effectiveness for intended national security applications.

[13] Report of the Defense Science Board on Task Force on DoD Supercomputing Needs, Oct. 11, 2000, Office of the Under Secretary of Defense for Acquisition and Technology

[14] *ibid*

Procurement of Operational Systems

Although not a part of the proposed IHEC development and acquisition plan, the procurement of commercial systems of course will be necessary to field the required national security capabilities and ensure the long-term viability of the domestic high-end computing industry. It is only through the procurement and operation of these systems that the benefits of additional HEC resources will translate into real products providing direct support of national security. In addition, such procurements are essential as they provide the industry the incentive necessary to ensure success of the IHEC Program. However, while the Committee recognizes the crucial importance of HEC system procurements, it feels that system procurements should be outside of the IHEC Program and remain under control of the individual Agency/Mission areas. This will provide agencies necessary autonomy and flexibility to procure the systems that are most appropriate to meet their mission requirements.

In exploring the procurement question, the Committee noted that the critical national security problems that rely on HEC fall clearly into two categories: (1) those that are limited mainly by the level of performance that U.S. industry is prepared to deliver according to current system roadmaps, and (2) those that are limited mainly by insufficient agency funding to procure currently planned systems. Cryptanalysis is an example of the former, while operational weather forecasting (e.g., the Navy's POPS Program) is an example of the latter. The IHEC Program is designed to address the first category. However, the committee concludes that the second category must also be addressed, with additional agency funding, to respond to the intent of the report accompanying the FY02 House Defense Appropriation.

It should be noted that the IHEC Program would be carrying out different initiatives in each of these phases simultaneously. Applied Research, Advanced Development, Engineering and Prototype Development, and Procurements will all be going on at the same time. This is shown graphically in Figure 2 (below).

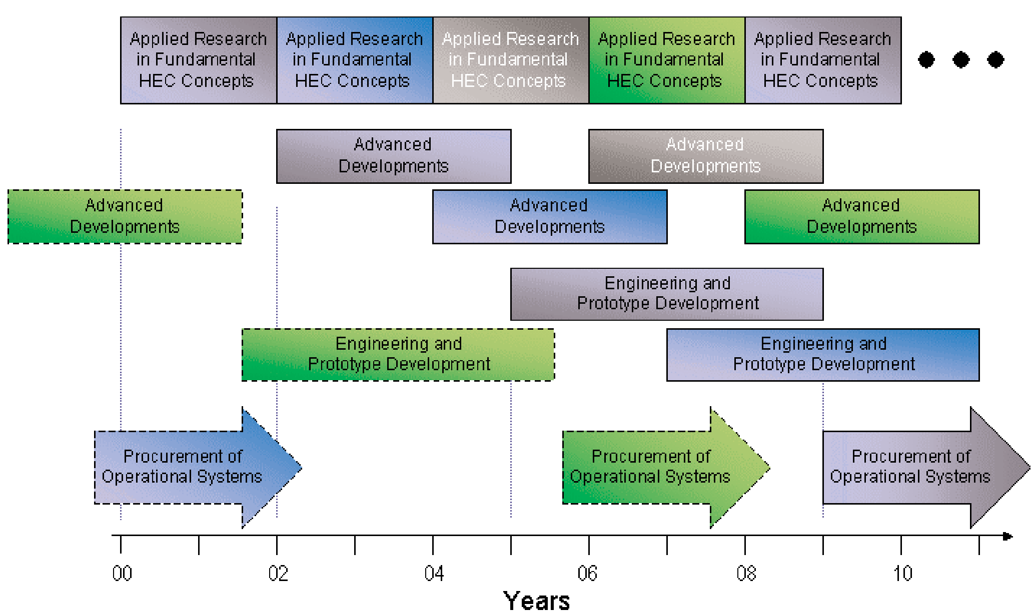


Figure 2 - Structure of the IHEC Program (Continuation of existing efforts outlined in broken lines.)

Next Generation Supercomputer Program Governance

The IHEC Program is needed and supported by the military services, NNSA, and multiple defense agencies. However, today each group independently focuses on individual mission requirements (with some loose coordination) and no single activity has the resources to support the long- and medium-term research and development efforts need to assure next generation supercomputers are designed and produced to meet our national security needs. While NNSA, DARPA and NSA all have modest research and engineering efforts in this area, they are not sufficient to create the necessary technology or maintain a healthy HEC industry. These existing efforts should be consolidated and expanded under the IHEC.

Because this effort supports such a broad community, a Joint Program Management Office (JPO) is the appropriate management structure for the Integrated High End Computing Program. The JPO would be structured to allow each of the involved agencies to actively participate in the selection of the efforts within the program. A recognized HEC expert would be recruited as its Director. Also, given the importance and scope of efforts described in this plan, a review board would be established with senior representatives from each of the involved organizations to oversee the program. The JPO, itself, would be staffed with members from the participating national security HEC community. As necessary, the JPO would also engage technical experts from the community it serves in collecting and reviewing critical mission needs as well as in building the overall R&E program. The JPO would then execute the IHEC program through HEC R&D agencies—NSA, DARPA, DOE/NNSA and NASA.

Furthermore, because of the high interest of the involved military services and defense agencies, the Joint Program Office would be created within the Office of the Director, Defense Research and Engineering (ODDR&E). A JPO at the ODDR&E level provides the appropriate structure to consolidate, champion, and budget for the needs of the collective community.

In addition to carrying out its IHEC program, focused on supporting technologies critical to national security, the program management office will be expected to interact closely with other government agencies that are active in the HEC field. For instance, the DOE/Office of Science, NOAA (National Oceanographic and Atmospheric Agency), and NIH (National Institutes of Health), have missions with substantial components that rely on modeling and simulation of complex systems, and the NSF has a broad science portfolio with simulation elements. The mission agencies address scientific problems essential to their missions and fund research focused on critical mission needs while NSF supports a wide range of activities important to the advancement of information technology as a whole. Mission agencies often establish large, nationwide, integrated programs aimed at solving the most challenging scientific and engineering problems. Many ongoing research activities by the mission agencies and NSF are either jointly funded or address difficult problems in a complementary fashion to the mutual benefit of the agencies and the nation. The program management office should therefore maintain close ties with these government agencies to maximize opportunities for mutually beneficial, collaborative research in the future.

Budgetary Requirements

The IHEC program funding has not been tied to any specific fiscal year. However, the panel recommends starting as soon as possible. Hence a concerted effort should be made to get the IHEC program included as part of the next DOD Program Objectives Memorandum (POM) with our recommendation for starting the IHEC program as part of the FY05 DOD POM.

Two funding levels were proposed: base and progressive. These are shown at their maturity levels (Program Year 5) in Table 1. The base-level program should be sufficient to keep the U.S. HEC industry engaged in addressing most of the HEC needs of the national security community. The progressive-level would allow additional higher payoff/greater risk efforts to be pursued. As quantum computing research is considered to be a separate activity, funding for it is not included in either program.

The IHEC program would be phased in over five years during which time its annual funding would be ramped up. Subsequent years would be funded at the Program Year 5 levels, adjusted only for inflation. This would be the case for both the Base-Level and Progressive-Level efforts.

The Base-Level Option will eventually grow to fund fifty projects in the Applied Research phase of the IHEC program. The profile for phasing-in the funding is given in Table 2. This is long-term applied research in fundamental HEC concepts. Funding will be spread out among academic institutions, start-up companies, as well as established large component and systems vendors who have a bright idea for a significant breakthrough. This phase would support a graduate-level scholarship program. Long–term research is also more likely to entail a high degree of risk and, hence, failure. This is to be accepted as necessary to push the frontiers of technology. Although the amount of funding per project could vary widely, $1M on average is reasonable. Each project could be expected to generally run for one to three years with renewals for promising technology available. The Progressive-Level Option extends the project base to one hundred simultaneous projects. Its profile for phasing-in funding is shown in Table 3.

The Advanced Development phase is more expensive, and will entail a smaller number of larger companies, as proofs of concepts are initially tested and early integration begun. Also, costs are much higher per project as higher levels of integration, testing, and materials are required. Each project can be expected to nominally cost $15M. The Base–Level Option funds four simultaneous projects while the Progressive–Level

Category	*Total Funding ($M)*	
	Base-Level	*Progressive-Level*
Applied Research	*$50*	*$100*
Advanced Development	*$60*	*$75*
Engineering & Prototypes	*$100*	*$150*
IHEC Labs & Infrastructure	*$40*	*$65*
Totals	*$250*	*$390*

Table 1 - Funding levels for the Base– and Progressive–Level IHEC Programs when fully implemented

funds five. As AD projects are more complex, a three-year cycle in the Advanced Development phase is appropriate.

The Engineering and Prototypes phase continues the trend of fewer projects, higher levels of integration, and correspondingly higher levels of cost. Projects in the Engineering and Prototypes phase will be executed exclusively by HEC vendors capable of bringing sophisticated modern supercomputers to market. The goal of this phase is to produce Serial Number 1 of a machine. Previous studies have shown that it takes approximately \$200M in non-recurring engineering (NRE) costs to bring a supercomputer to market[15]. Assuming a four–year cycle, the annual costs per system are \$50M. The Base–Level Option funds two simultaneous projects while the Progressive–Level Option funds three. These Engineering and Prototype efforts will be phased in as shown in Tables 2 and 3.

Finally, operating the IHEC Laboratories as well as execution of the program itself needs to be funded. It was estimated that this would add approximately another 20% to the cost of the program.

Program Year	Funding ($M)				
	Applied Research	Advanced Developmemt	Engineering & Prototype Development	IHEC Labs & Infrastructure	Totals
1	\$40	\$60	\$50	\$30	\$180
2	\$45	\$60	\$100	\$40	\$245
3	\$50	\$60	\$100	\$40	\$250
4	\$50	\$60	\$100	\$40	\$250
5	\$50	\$60	\$100	\$40	\$250

Table 2—Funding profile for Base-Level IHEC Program

[15]This number has held remarkably consistent dating back to the Control Data Corporation 7600.

Program Year	Funding ($M)				
	Applied Research	Advanced Developmemt	Engineering & Prototypes	IHEC Labs & Infrastructure	Totals
1	$40	$60	$50	$30	$180
2	$60	$75	$100	$45	$280
3	$80	$75	$100	$50	$305
4	$100	$75	$100	$55	$330
5	$100	$75	$150	$65	$390

Table 3—Funding profile for Progressive–Level IHEC Program

Conclusion

The IHEC program was prepared at the request of Congress to ensure that current and future HEC needs of the national security community are addressed. It incorporates the best ideas of the leading experts in the HEC field from government, national laboratories, industry and academia. NSA, in cooperation with DARPA, DoD HPC Mod Office, NRO, DOE/NNSA, and NASA drafted this report. It describes the need for the IHEC program, assesses the current and future state of the HEC industry, and presents the salient features, including budgetary requirements, of the IHEC program.

OUWG Summary

High Performance Computing, including advances in both software and hardware, has revolutionized the manner in which the national security community addresses a wide variety of critical national challenges. Many defense programs that give the United States dominance over its adversaries could not exist in their present state without a strong domestic commercial source for High Performance Computing hardware and new developments in computational science.

This report has identified 10 defense applications[16] that rely on High Performance Computing for both current and future needs:

1. Comprehensive Aerospace Vehicle Design: is the simulation and modeling of advanced fighters, cruise missiles, and reconnaissance aircraft extending into hypersonic velocities.

2. Signals Intelligence: is the transformation, cryptanalysis, and intelligence analysis of foreign communications on the intentions and actions of foreign governments, militaries, espionage, sabotage, assassinations, or international terrorism.

3. Operational Weather/Ocean Forecasting: provides worldwide, 24-hour weather guidance to the military, CIA, and Presidential Support Unit for current operations, weapons of mass destruction contingency planning, etc.

4. Stealthy Ship Design: future Naval ships must accommodate emerging technologies such as electric propulsion systems and stealth to reduce detection signatures.

5. Nuclear Weapons Stockpile Stewardship: in the absence of nuclear testing, stockpile certification will be based on the development and validation of high quality physics simulations of nuclear weapons.

6. Signal and Image Processing: to find high value targets requires the analysis of multispectral signals and images and the processing of more and improved sensor data at higher resolution in less time.

7. Army Future Combat Systems: provides for rapid response and deployment of forces using advanced logistical analysis, as well as the discovery, refinement, development, and evaluation of critical lethality and survivability technologies.

8. Electromagnetic Weapons Development: includes the development of the airborne laser for ballistic missile defense and other electromagnetic devices to disable battlefield electronics and provide alternatives to lethal weaponry.

9. Geospatial Intelligence: the analysis and visual representation of security-related activities in reference to the Earth.

10. Threat Weapon Systems Characterization: includes the development and application of advanced modeling, computational analysis, and simulation methodologies to the study and characterization of foreign threat weapon systems, as well as the prediction of the performance of those threat systems in complex operating environments.

Defense applications identified key computational characteristics as being a bottleneck. By far, the number one priority is for improved memory subsystem performance to include memory latency, memory bandwidth, and memory size. Other commonly identified computational characteristics[17] inhibiting performance include: more

[16] This is not an exhaustive list; time limitations constrained the scope of the study. For example: chemical, biological, and radiological attack modeling in real time, and critical infrastructure protection are but two examples of emerging applications that are not yet mature enough for well-specified requirements.

[17] Signals Intelligence also has severe time–to–solution requirements.

productive programming environments, better CPU performance[18], and I/O subsystem performance. Most of the requirements described in this report, with the exception of the signals intelligence, can generally be accomplished with commercially developed hardware. However, all the applications described here in would benefit from specialized systems addressing the above-mentioned bottlenecks.

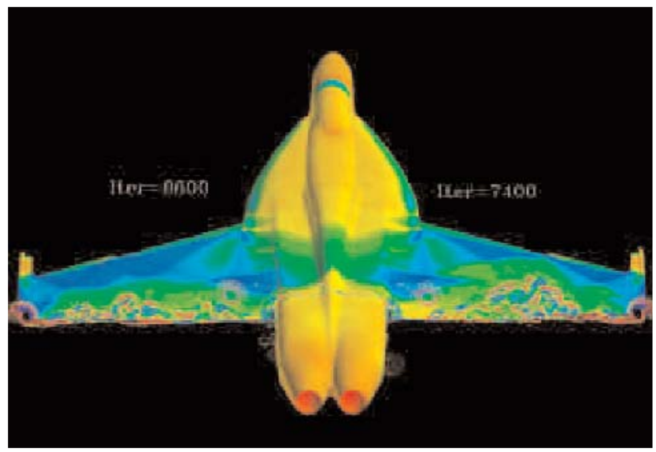
Summary of Key HPC Work

High Performance Computing, including advances in both software and hardware, has revolutionized the manner in which the national security community addresses a wide variety of critical national challenges. State-of-the-art high performance computing systems are used to decode the communications of those who would harm our nation, design new generations of combat aircraft, forecast global weather conditions, safeguard our nuclear stockpile, and analyze vast amounts of information essential to our homeland defense. Our ability to effectively utilize HPC systems is growing at an exponential rate. Summarized below are several key application areas that highlight the national security importance of a strong domestic HPC industry.

Comprehensive Aerospace Vehicle Design

Global reach and global dominance is critically dependent upon air vehicles. All three military services and NASA are making substantial investments to develop future aerospace vehicles that will extend U.S. dominance in the coming decades. High performance computing is playing an increasingly vital role in current programs, such as the F-22, V-22, and JSF. And, HPC will play an even greater role as we look to future aerospace development programs involving hypersonic (speeds between 5 and 30 times the speed of sound) capabilities such as the National Aerospace Initiative. The design of these weapon and space transportation systems requires more comprehensive physics models in order to accurately simulate these harsh flight regimes.

Design of aerospace vehicles is of necessity a multi-disciplinary effort. Current capabilities to model the external airflow, propulsion performance, vehicle signature, and materials properties allow reasonable predictive results on today's HPC systems when computed independently. One of the primary drivers for exponentially growing HPC requirements is the need to combine these independent modeling efforts into an interactive modeling capability that allow for interaction among the model components. For example, coupled models will allow engineers to quickly see the effect of proposed changes in the propulsion design on the vehicle stealth signature.



The National Aerospace Initiative develops and demonstrates technologies for hypersonic systems. For DoD in the near term, hypersonic systems include hypersonic cruise missiles and in the far term include hypersonic strike or reconnaissance aircraft and affordable on-demand access to space vehicles. Hypersonic cruise missiles would prosecute time critical targets, flying hundreds of nautical miles in minutes with cruise speeds of

[18] Most applications require floating point performance. Signals Intelligence was unique in their requirement for significantly enhanced integer and logical CPU perform ance.

Mach 6 to 8. Hypersonic strike or reconnaissance aircraft would provide tactical commanders with near real-time reconnaissance information on demand. Cruise speeds would be between Mach 5 and 8. Hypersonic-powered space vehicles would enable true on-demand access to space to meet military operations tempos. Computational techniques are required to assess scramjet engine operability from Mach 4 through Mach 8 in the near term and from Mach 10 through Mach 15 in the far term. These efforts will enable efficient development and demonstration of scramjet engines and then combining the scramjet with rockets and or high-speed turbine engines capable of operating efficiently over the full range of Mach numbers.

Signals Intelligence

The Signals Intelligence mission is to intercept and analyze foreign adversaries' communications signals, many of which are protected by codes and other complex countermeasures. DoD must collect, process, and disseminate intelligence reports on foreign intelligence targets in response to intelligence requirements set at the highest levels of government. The Signals Intelligence mission targets capabilities, intentions, and activities of foreign powers, organizations, or persons. It also plays an important counterintelligence role to protect against espionage, sabotage, or assassinations conducted for or on behalf of foreign powers, organizations or persons, or international terrorist groups or activities.

There are two main users of HPC within the Signals Intelligence mission: Intelligence Processing and Intelligence Analysis. Intelligence Processing seeks to transform intercepted communications signals into a form that can be understood by humans. This may entail overcoming sophisticated cryptographic systems, advanced signals processing, message reconstruction in the presence of partial or corrupted data, or other complex signaling or communications subsystems. Intelligence Analysis begins once the communications of interest have been transformed into a human readable form. A blizzard of communications messages exist; the challenge is to correlate these apparently unrelated transmissions into a complete mosaic so that knowledge of adversaries' intentions can be discerned and actionable intelligence provided to the war fighter or National Leadership.

The HPC requirements of Signals Intelligence are without bound. Cryptographic systems are deliberately designed to withstand the most intensive assaults of the most advanced computing systems. Furthermore the increasing volume, variety, and velocity of data presented to the Signals Intelligence system present enormous challenges to the agencies charged with the defense of the United States. Finally, Signals Intelligence has severe time-to-solution requirements; intelligence delivered to the National Security Community after an event has occurred is of no value to the nation.

There are three main HPC requirements of most significant importance to Signals Intelligence. First is improving memory performance. Processor performance is increasing at a rate far in advance of the memory's ability to sustain peak processor utilization. This is especially true for algorithms exhibiting significant percentages of random memory access patterns. Of secondary importance is the whole area of "ease of use". As modern day distributed memory and parallel processing systems become more ubiquitous and more complex, the number of programmers who can quickly develop software to take full advantage of those systems continues to decline. Hardware utilization, programmer efficiency, and time-to-solution need to be dramatically improved. Finally, algorithms needed to prosecute the Signals Intelligence mission make intensive use of the integer and logical performance of processor design, and have less demanding requirements of floating point than many other applications. This requirement seems to be unique within the DoD community.

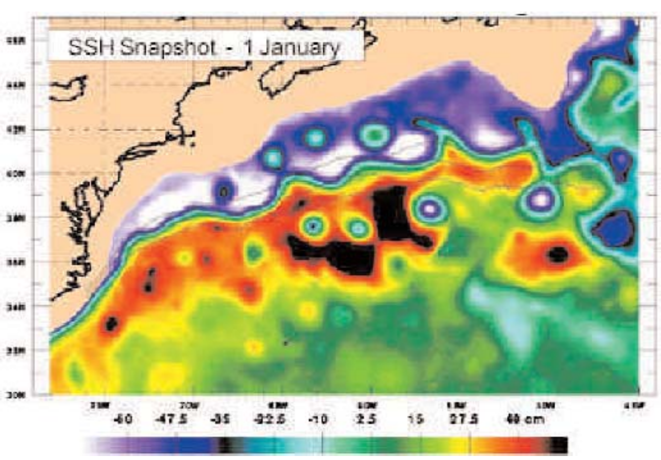
Operational Weather/Ocean Forecasting

The Primary Oceanographic Prediction System (POPS) hosted at the Fleet Numerical Meteorology and Oceanography Center (FNMOC) produces and provides critical, classified and unclassified atmospheric and oceanographic guidance to Department of Defense (DoD) activities worldwide on a fixed schedule, 24 hours a day. POPS is the engine that drives the <u>entire</u> Navy's global, regional, tactical atmospheric, oceanographic, wave, ice, and tropical cyclone models; is DoD's only coupled air/ocean model; and is <u>the only national system</u> that assimilates classified and unclassified data, and produces and disseminates classified and unclassified global/regional atmospheric guidance that is used by:

The Navy to operate their global ocean, regional, and tactical ocean, atmosphere, wave, ice, tropical cyclone models from the unclassified to the SCI level, and their eight distributed tactical forecast systems;

The Air Force to operate their regional atmospheric models, cloud prediction systems, and strategic decision aids as specified via the Navy/Air Force agreement;

The Joint Forces Command, Defense Threat Reduction Agency, and Lawrence Livermore National Laboratory to operate their Weapons of Mass Destruction decision aids and to aid in contingency planning;



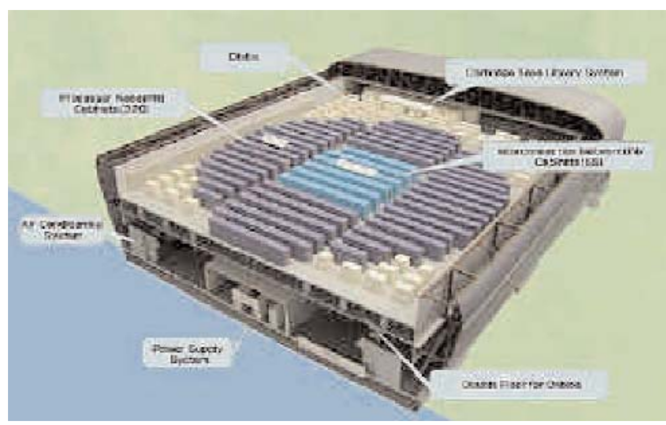
The Central Intelligence Agency to aid in their contingency planning;

The national backup to the National Weather Service's supercomputer. There is no global backup for the nation;

The U.S. STRATCOM for ballistic missile support; and

The Presidential Support Unit to aid in their contingency planning.

As evidenced by Japan's recent $350M dollar earth simulator investment, the United States is falling behind the industrialized world in weather prediction capability. A lack of enhanced capabilities will have a ripple effect within a number of areas in the Navy and DoD and will precipitate a crisis due to the lack of state-of-the-practice weather/ocean prediction systems. Inadequate



high-end operational supercomputing resources in DoD have hampered the ability to continue providing real-time weather and coupled modeling forecasts to the warfighter. The POPS will soon be outdated, leaving our forces with an inferior capability to predict weather in various operating environments. This is especially critical during a time when Enduring Freedom, Homeland Security, and the War on Terrorism are placing a demand for very-high resolution atmospheric and oceanographic models that are able to calculate the dispersion of air/water-borne threats to the

Fleet and the nation. POPS will not be able to provide tailored high-resolution oceanographic and atmospheric prediction and dispersion products that are needed to address the full range of threats facing the nation daily. The POPS will not be able to operate the new weather models that are scheduled for transition in the FY 04–FY 09 timeframe, including: aerosol/chemical dispersion forecasts, target area weather predictions, support to on-scene modeling, improved tropical cyclone forecasts, and very-high atmosphere/space forecasts.

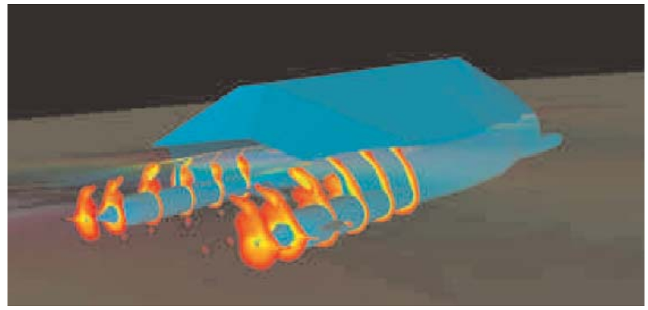
The science and technology community is developing appropriate models and simulations for the future and these capabilities can be implemented with appropriate HPC investments (development & capital investments). Additional information can be found in the Administrative Model Oversight Panel (AMOP) modeling Roadmap:

https://www.cnmoc.navy.mil/nmosw/staff/roadmap/cnmocweb/web-pages/cnmochome.html.

Stealthy Ship Design

Naval ships of the future must be capable of accomplishing emerging operational requirements such as increased littoral operations, and accommodate emerging technologies such as electric propulsion systems and stealth. To meet these requirements, future ships will be radically different from those of today. Current design and analysis methodologies are primarily based on potential-flow theory and empiricism and cannot adequately design or predict future hydrodynamic ship signatures. The U.S. Navy has turned to high performance computing to design the next generation of surface combatants.

For example, the DD-X land-attack surface combatant requires ship signature reduction to levels comparable to submarine signatures. Providing the new computational capability will



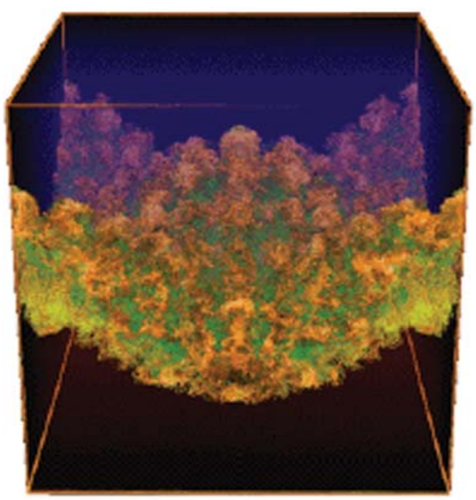
significantly reduce the developmental cost and design-cycle time for the U.S. Navy's future stealthy ships, and improve the chances of success. Without this capability, signature mitigation requires a traditional build-and-test approach that could prove to be prohibitively costly and time-consuming.

Current prediction capability needs to be extended to coupled 6-DOF motions in seaways and to accurately simulate the turbulent wakes, steep

breaking waves, air entrainment, and the generation of spray for a full scale surface combatant. A series of validations from model– to large– to full–scale needs to be made to assess the accuracy of the modeling of the complex flow physics, thus improving the prediction capability. Such calculations are computationally intensive because of the turbulent nature of the flow and because of the fidelity necessary to resolve complex free-surface interactions. Additionally, such computations need to be done much faster than is feasible today in order to do design trade-offs and hull form optimization in an acceptable time frame to impact design decisions. To achieve the above goals, the hardware should improve in speed on the order of 20–30 times the current capabilities

Nuclear Weapons Stockpile Stewardship

The Nuclear Weapons Stockpile Stewardship Program oversees the certification of the nuclear stockpile. In the absence of nuclear testing, stockpile certification will be based on expert analysis using validated computational simulations of the performance of nuclear weapons together with a strong above ground experimental program. The high performance computing portion of the project is the development and validation of high quality computational physics simulations of nuclear weapons and the use of those simulations for the analysis of nuclear weapons systems in support of the stockpile certification. Without underground nuclear tests, the computer simulations must have a vastly improved predictive capability based, to the greatest extent possible, on accurate physical data.

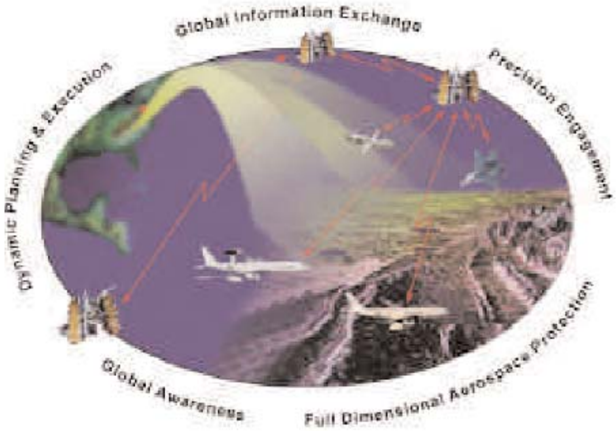


Without that confidence, the U.S. military strength will be substantially reduced. The U.S. will then be much more vulnerable to hostile powers and the security of the U.S. will be substantially reduced. In particular, we will be more vulnerable to foreign powers that possess nuclear weapons. Achieving the desired increase in predictive capability will require substantial improvements in spatial, energy, and temporal resolution, better mathematical algorithms and more accurate physical data. Achieving the required improvements will require an increase in computer capability of roughly 10^5. High performance computing is the only way to achieve that required increase.

The nuclear weapons codes being developed as part of the Accelerated Strategic Computing Initiative are complex multi-physics codes. The codes integrate initial value partial differential equations for the conservation of particles, momentum, and energy for the important element and constituents of nuclear weapons. Typical calculations use 10,000 to 1,000,000,000 mesh cells and require thousands of processors depending on the problem, desired resolution, and complexity of the nuclear package. The most recent calculation took 140 processor days to complete. A further ten-fold increase in resolution coupled with a 10–100-fold increase in computational capability, as higher-fidelity physics is added to simulations, is required. Time to solution must be reduced from 140 days to 7 days in order to allow sufficient parametric analysis. These codes create massive data files that necessitate terabytes of secondary storage available with very broad I/O pathways to prevent delays in calculations.

Signal and Image Processing

More than ever before, the ability to effectively manage battlefield and intelligence information is essential to our nation’s defense. Our ability to quickly find high value targets and decisively target the right weapons will make the critical difference in future conflicts. Multi-spectral signals and images collected by national and tactical assets must be quickly analyzed to determine possible threats associated with appropriate and available weapons systems, and then converted into appropriate intelligence and/or target acquisition products. High performance computing is essential to process more and improved sensor data at higher resolution in less time. Improvements in data



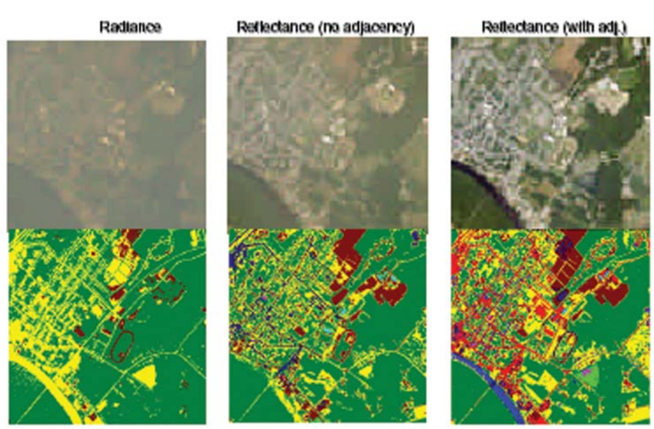
gathering and collection systems, in space and on the ground, creates a tremendous opportunity that can only be exploited by increasing the processing systems giving them the flexibility to accommodate tasking with priority and preemption.

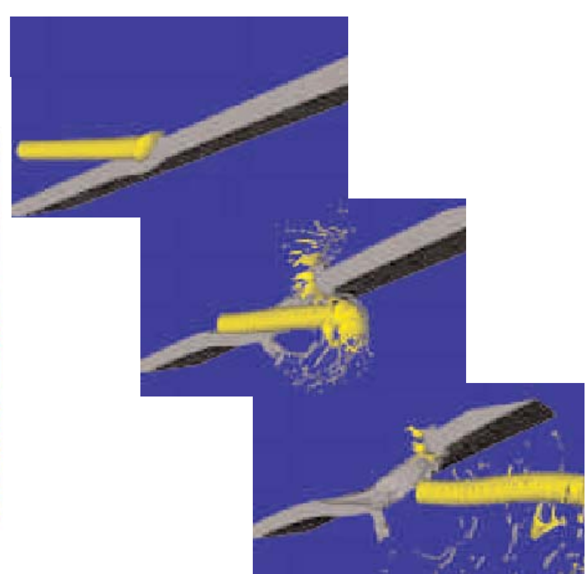
Improvements of several orders of magnitude in processing power with well balanced architectures capable of handling, processing, disseminating streams of 'big data'…. capable of moving large continuous flows both within architected processing system and continuous real-time ingress from sensors and platforms and egress to projectors/displays of data for knowledge-based situation awareness are needed.

Army Future Combat Systems

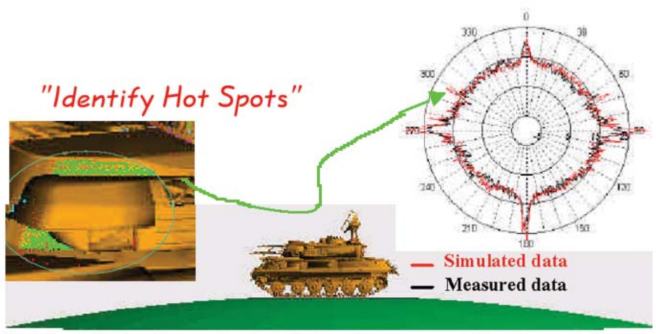
Today's uncertain world requires rapid response and deployment of decisive ground forces. The Army is transforming their heavy forces to be more strategically responsive and upgrading their light forces to be more lethal and survivable. Strategic responsiveness means deploying, anywhere in the world: a brigade in 96 hours, a division in 120 hours, and five divisions in 30 days. In order to achieve these deployment timeframes, a Brigade Combat Team must fit all materiel required for deployment on a C-130 cargo aircraft. This means that each combat system or combat platform must weigh less than 20 tons and be tailored to achieve the ground combat and mobility requirements essential to battlefield dominance. The Army's science and technology (S&T) community is leading the way for long-term transformation through large-scale simulations.

The U.S. Army Research Laboratory (ARL) is focused on the development and evaluation of lethality and survivability technologies germane to future land combat systems (e.g., FCS). ARL's research programs leverage high performance computing (HPC) to solve problems associated with the discovery, refinement, development, and evaluation of critical lethality and survivability





technologies. Examples include low observable survivability, low/high frequency RF spectrum, IR, coatings, and acoustics in the presence of the battlefield environment (including terrain and weather effects), advanced kinetic energy (KE) penetrators, multi-functional warheads (blast, fragmenting, shaped charge, explosively formed penetrator, and others), kinetic energy missiles, passive armors, reactive armors, and other advanced/hybrid armors. The penetration mechanics of advanced KE penetrators, as well as novel penetrator concepts, impacting modern and emerging armors, is complex. Only a portion of the dynamics of penetrator-target interaction can be gleaned solely from terminal ballistic experiments. A large portion of these dynamics can only be examined with modeling and simulation.



Modeling Radar Cross Section in the Presence of the Ground Plane

Achieving the Army's transformation goals for an objective force that is strategically responsive and dominant across the full spectrum of operations will enable future Army operations; from small-scale contingencies including anti-terrorism operations to major theater wars. Advances in information, materiel, and weapons systems technologies will make it possible for objective force units to achieve at least the same effectiveness as today's forces, but with fewer, lighter, and more sustainable systems.

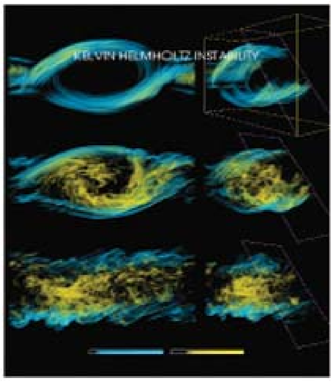
Electromagnetic Weapons Development

Use of electromagnetic energy in a weapons device has been the dream of many war planners for decades. That dream is about to become a reality with the development of the airborne laser (ABL) for ballistic missile defense. In addition, the Air Force is investigating the use of other electromagnetic devices to disable battlefield electronics and provide alternatives to lethal weaponry in

denying enemy access to specified areas. Nonlethal weapons are of special interest for anti-terrorist and urban operations that may involve combatants mixed with civilian populations.

The ABL represents the boost-phase component of a layered U.S. ballistic missile defense system. Chemical lasers provide the ability to impact a target at the speed of light and at large ranges; this provides the ability to hit and kill intercontinental and theater range ballistic missiles, satellites, supersonic cruise missiles and a variety of other targets. ABL is designed to destroy enemy missiles in the first 3 to 5 minutes after launch, and in so doing, address multiple ballistic missile systems with different operational ranges. By destroying a missile in its boost phase, the ABL will scatter debris from a successfully engaged missile over enemy territory so that it does not threaten the missile's intended target area.

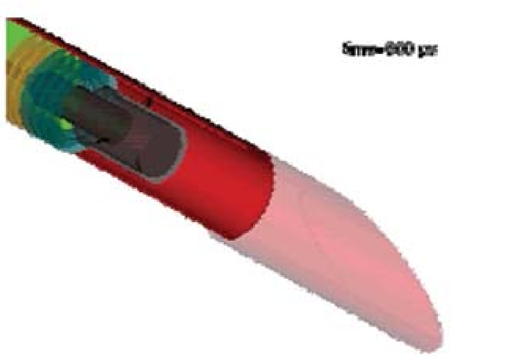
ABL system design and performance evaluation is critically dependent on the capability to compute the performance of the laser and the



effects of turbulence on optical propagation of the laser and adaptive optics used to position it. Reduced ABL effectiveness and/or non-optimal deployment will result if we do not properly characterize the impact of stratified atmospheric turbulence on optical propagation. In addition, once the impact of turbulence at operational altitudes is understood, the ability to forecast turbulence strength will be an important operational asset.

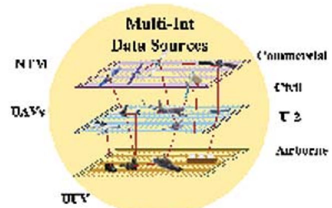
High Performance Computing is enabling the design and test of the ABL. First, laser power and gain are being predicted using simulations of the chemical laser system. Second, the strength, distribution, and spatial spectrum of stratospheric turbulence are being modeled and simulated. And third, optical propagation phenomena and the effects of stratospheric turbulence are being modeled to predict performance and assist in the design of the optical tracking and adaptive optics systems.

Given the importance of electronics on the modern battlefield, the ability to target these electronics will give the U.S. a distinct advantage against technically advanced adversaries. The effect of RF radiation on the electronics can vary from transient disruption to destruction of the hardware, depending on the power and frequency of the radiation. Furthermore, since only electronic equipment is affected, RF radiation can be applied in a non-lethal manner, giving the war fighter a broader range of options for dealing with opponents of varying sophistication and strength. Several different kinds of microwave devices are being evaluated.



Geospatial Intelligence

Geospatial intelligence comprises key reference knowledge for all planning and operational national security missions. The ability to acquire, analyze, assimilate, and distribute intelligence from multi-source, multi-purpose applications requires an understanding of the geospatial and temporal context of the data, and involves a computational-intense computing capability.



Today's computational approaches allow "data spillage" to occur in the intelligence derivation process, whereby some intelligence may be lost as a result. The intelligence loss problem is exacerbated with the introduction of motion imagery as a new data source for accomplishing missions such as pattern recognition and moving target tagging and tracking. This prevailing condition (loss of possible strategic and actionable intelligence) is high risk in our business, but is where high-end computing can offer valuable risk reduction.

Areas where the need for greater computational capability increases include both internal multi-source data fusion applications as well as multi-source ingest from multiple collection systems, including those in open source and commercial-based origins. The necessity to move toward multi-int processing, while retaining geospatial and temporal “ground truth”, will further complicate the ability to integrate or fuse massive data and support its visualization effectively. The processing demands on computational applications to support these needs are increasing in the areas of on-board sensor data preprocessing, ground or receiver-based data processing, analysis, data management, information rendering, and visualization for multi-purpose usage in real-time, near real time, and strategic applications. These applications vary in scale from the very large, such as characterizing global regional transportation networks to the very small, such as finding toxic indicators in plant cells on a specific roadway. Yet in each case, a suite of significant computational transactions are involved.

There is a need for increased autonomous, and peer-to-peer processing for geospatial intelligence to support real-time strategic and tactical analysis. The future of military operations will involve increased needs for sensor to weapon location and precision guidance, advanced robotic sensing, geospatial condition understanding, unmanned vehicle control, and battle space environment characterization for mission planning and execution. The concept of building geospatial intelligence for these future missions through new analytical visualization constructs that support full emersion in a multi-source, real-time context, will require high end computing research and development.

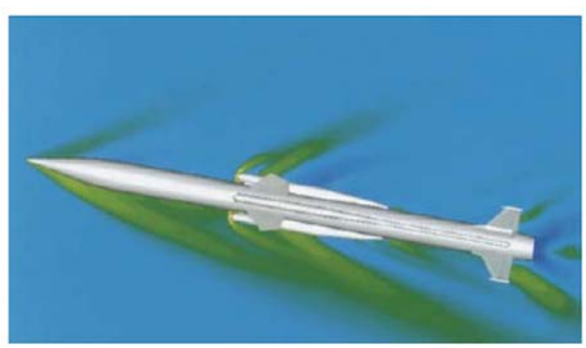
There is a need for the geospatial intelligence community to understand impact and usage of quantum computing applications in research. This research requires high-end resources. The need follows the trend to build data preprocessing into advanced sensors for immediate operational or analysis usage (e.g., upstream, preprocessing of images, and data on the satellite). Establishing geospatial and temporal ground truth fundamentals in quantum applications is a growing need in signals and image processing methods.

Threat Weapon Systems Characterization

Accurate characterization of threat weapon systems is an essential element of national security, whether viewed from military operations or homeland defense. The Defense Intelligence Agency has established and maintains its High Performance Computer System (HPCS) at the Missile and Space Intelligence Center on Redstone Arsenal, Alabama, as a Defense Intelligence Community asset for all-source analysis and production of intelligence.

Based upon a variety of tightly networked architectures, the HPCS is configured each year to address key scientific and technical intelligence issues. Both the software and hardware configurations are adjusted annually to maximize the ratio of computing power to operating costs. Currently the focus of its operations is to support computational aerodynamics, signature prediction, and missile performance simulation activities – all in direct support of intelligence customers such as the Missile Defense Agency, the Combative Commands, and also against issues related to homeland defense.

Faced with characterizing literally hundreds of often poorly defined threat missiles and other



weapons systems, MSIC and the rest of the Defense Intelligence Community depend upon high performance computing available from the HPCS to estimate weapon characteristics. As the sophistication of the threat systems increases, the requirements for precise, accurate threat estimates also increase dramatically. The capabilities of the HPCS, therefore, must continually improve and expand. Otherwise, the threat information required to insure the success of US interests and forces, will not be available.

SUMMARY OF OUWG TECHNICAL REQUIREMENTS

The computational requirements of the national security community are substantial. In terms of "raw" computing power, over 7,279 TeraOp years[19] are needed by 2012. Current HPC procurement programs are not funded at levels needed to provide this level of computing capability.

Project	Number of Sustained Giga op/s Required Each Fiscal Year FY 2003 (GO-Yrs)	FY 2005 (GO-Yrs)	FY 2007 (GO-Yrs)	FY 2010 (GO-Yrs)	FY 2012 (GO-Yrs)
Total DoD Requirements	77,160	139,016	279,875	1,632,231	5,288,430
Signals Intelligence	Classified	Classified	Classified	Classified	Classified
Nuclear Weapons Stockpile Stewardship	74,600	256,100	415,000	835,000	1,991,000
Total HPC Requirement (Less NSA)	151,760	395,116	694,875	2,467,231	7,279,430

All of the computational projects described in the previous section cannot be effectively computed on commercial supercomputers driven by the personal and server computer markets. The developers and users of these national security applications identified the key computational characteristics that they believe are the major bottlenecks to improved productivity. By far, the number one priority is for improved memory subsystem performance to include memory latency, memory bandwidth, and memory size. Other commonly identified computational characteristics[20] inhibiting performance include: more productive programming environments, better CPU performance[21], and I/O subsystem performance. The following table summarizes these key findings. Much more specific performance information is provided on each key application area in Appendix C.

[19] A TeraOp per second is 1 trillion (10^{12}) integer or floating point operations per second. A TeraOp year is 1 trillion (10^{12}) integer or floating point operations times the number of seconds in a year (31.5M seconds) or 31.5 x 10^{19} computer operations.

[20] Signals Intelligence also has severe time–to–solution requirements.

[21] Most applications require floating point performance. Signals Intelligence was unique in their requirement for significantly enhanced integer and logical CPU performance.

Bottleneck	Identified by
Memory Performance (Latency/ Bandwidth/Size)	Nuclear Weapons Stockpile Stewardship, Signals Intelligence, Army Future Combat Systems, Operational Weather/Ocean Forecasting, Electromagnetics Weapons Development
CPU Performance	Army Future Combat Systems, Operational Weather/Ocean Forecasting, Stealthy Ship Design, Comprehensive Aerospace Vehicle Design, Electromagnetics Weapons Development
More Productive Programming	Signals Intelligence, Comprehensive Aerospace Vehicle Design
I/O System Performance	Operational Weather/Ocean Forecasting, Environments Electromagnetics Weapons Development

The needs of the national security community are large and varied as evidenced by the subset of application areas described in this report. The current commercially available products are effective at meeting a significant subset of today's needs but the server dominated markets are forcing the major vendors to make design trade-offs that will not address the full scope of our growing future needs.

SYSTEMS, ARCHITECTURE, PROGRAMMABILITY, AND COMPONENTS WORKING GROUP

Findings

The Systems, Architecture, Programmability, and Components Working Group (SAPCWG) consists of high performance computing experts from the federal national security mission agencies, national laboratories, academia, and industry. The key findings of the SAPCWG are summarized below, followed by a set of recommendations.

TWO CLASSES OF SUPERCOMPUTERS

There are two classes of supercomputers with quite different outlooks. There are important National Defense applications requiring each type. The first category, sometimes called type-T machines, is built out of large production volume microprocessors. They are very efficient machines for applications with a high ratio of computation to communication. Because of the substantial leverage of components built for commercial data processing possible, several United States companies will continue to dominate this segment. The second category, which has been termed type-C, features high communications bandwidth to global memory, and between processors. The CPUs are typically developed for supercomputer application and are consequently sold in low volumes. This class of machine is required for applications with significant rates of random memory access and low ratios of computation to communications. Absent government investment, a domestic source of type-C computers is unlikely.

QUANTUM COMPUTING BALANCE

Quantum computers (QC) will not displace conventional high performance computers. There are only a very few problems for which theoretical QC algorithms are known to produce exponential speedup. If feasible, and barring unforeseen breakthroughs, QC of the scale necessary to attack even those limited problems at a level useful for national security is at least 20 years away. QC should therefore be considered outside the scope of this program.

U.S. government funding for research and development in technologies aimed at high-end computing has remained nearly constant at $50-70 million annually over the past five years. However, the portion of this investment spent on Quantum Computing has risen from just over zero to $60 million, with the result that only about $10 million of non-QC long lead-time HPC research is now underway in U.S. Universities and laboratories. This situation is out of balance.

IMPORTANCE OF EASE OF USE

Today's high performance computers are much too hard to use (100 times more difficult/expensive to program than desktop computers).[22] The result is that "hero" programmers capable of using the big machines become the scarce resource and hence the National Defense is underserved.

GAP IN TECHNOLOGY TRANSFER

There is a technology transfer gap. The U.S. computer industry is very good at engineering computers out of proven technology. In order to shepherd innovative technologies and architectures into production, the government must shoulder some of the risk by implementing a comprehensive long-range high end computing R&D program, the central core of which will sponsor advanced developments and engineering prototypes.

To address the findings summarized above, the SAPCWG group makes the following recommendations:

[22] C. Pancake, "Usable HPC - An Oxymoron?" Proceedings from the Conference on High Speed Computing, LANL, 1999.

INVESTMENT REQUIRED

Long lead-time applied technology research needs investment, initially returning to the pre-QC levels of five years ago, and ultimately substantially higher. This investment should be applied to the following areas of research: System Architectures; Memory Subsystems; System Software (including Parallel Languages and Programming Environments, Parallel Run-time and Operating Systems, Parallel Program Development Tools); Packaging/Power/Thermal Management; Interconnects and Switches; Storage and I/O; Novel Computation Technologies.

ADVANCED DEVELOPMENT

As new technologies advance, the DoD should fund a number of advanced development projects to build and test components and novel architectures for HPC systems.

PROTOTYPES

The DoD should undertake to fund the Engineering and Prototype Development of next generation supercomputers. These should be selected by a competition between available novel architectures and should be executed primarily by industry in collaboration with researchers. As there are national security mission needs for more than one type of supercomputer, more than one such development may be run concurrently.

TIME-TO-SOLUTION

The above research should focus on a unifying metric of Time-To-Solution (TTS). TTS considers program development, setup, and execution time, as appropriate for the individual application. TTS is influenced by system architecture, programming environment, component characteristics, and appropriateness of the system to the application.

Introduction to the SAPCWG Report

High Performance Computing (HPC) is the use of conventional and unconventional computing technologies and supporting infrastructure to solve difficult, computationally intensive scientific, mathematical, or engineering problems. Supercomputers are the high end of high performance computing. A supercomputer is sometimes (somewhat arbitrarily) defined as one of the 30 most powerful machines on the planet, or (equally arbitrarily) as a computer within a factor of 10 of the most powerful machine in the world by some measure. While these definitions are not particularly "definitive" or even useful, they do highlight the fact that no one characteristic defines what it means to be a supercomputer. Individual computer designs make tradeoffs between (among other things) instruction rate, number of processors, communications bandwidth between processors, bandwidth to memory, cache sizes, memory sizes, input/output rates, storage, programming model, and cost. The right mix of tradeoffs depends critically on the applications to be run; the "best" machine for one application may not even be capable of running another.

The SAPCWG committee assembled input from leading scientists from US Department of Defense mission agencies, Department of Energy National Laboratories, academia, and United States computer vendors to produce this baseline high end computing (HEC) forecast. This working group then used the results of the baseline to outline the technical characteristics of an Integrated HEC Program.

Baseline Forecast

SYSTEMS

The supercomputer market today is supplying two different classes of machines resulting from two different approaches to building high capability systems. The most broadly utilized approach, clustered symmetric multi-processors (SMP), uses multiprocessor nodes built with the same high-end, high volume microprocessors used in servers and desktop machines. This type of machine has been termed type-T because of the emphasis on transistors (the computational logic components within the microprocessors). The other approach, by contrast, has been labeled type-C because it concentrates on communications (both between processors and between processors and memory). This class of machine tends to use low volume, supercomputer intended, processors. Many other components of such systems are customized as well. Hybrids of the two approaches are also possible. The SAP Working Group finds a role for both types of architectures, as well as hybrids, but finds significant differences in the outlook for the two types of architecture.

The most prevalent type-T approach to building supercomputers is to cluster together a large number of symmetric multiprocessor nodes with either an ad-hoc interconnection network (e.g. Ethernet), or a proprietary one designed for a specific architecture. The nodes themselves are based on commodity microprocessors and memory parts designed and built for the $80B server and workstation market. The memory systems rely on multi-level caching strategies for performance, providing relatively low numbers of global memory accesses per peak floating point operation (in the range of $1/10^{th}$ of a byte per 64 bit floating point operation) for applications that exhibit sufficient locality of reference. Additionally, 2-way SMPs on one die are becoming commercially available (the IBM Power among others). Thus semi conductor technology is hastening the trend towards clustered SMPs. The architecture of the nodes is primarily driven by commercial applications such as transaction processing and information retrieval, which can benefit from the high cache bandwidth or the high clock speed provided by microprocessors. The vendors building this class of machines cannot be budged from their focus on supplying products for this large market (small nudges are possible but difficult). The promise of the clustered SMP approach is that by leveraging building blocks that are subsidized by the much larger high capacity server market, the relatively smaller ($1.2B) high end capability computer market can share in the economies of scale and therefore achieve performance that would otherwise be impossibly expensive. U.S. industry has done a very good job of developing microprocessor technology, in fact, quickly closing the gap in raw peak performance with respect to type-C processors. The same cannot be said for the global memory systems, which remains a serious impediment to HEC (High End Computing).

As a result, SMP clusters can deliver high peak operations per dollar, especially for applications that can take advantage of memory caches. However, machines of this type, when applied in the context of a massively parallel supercomputer can only achieve between 1% and 10% of their peak processing potential on many important application codes which exhibit higher ratios of communication to computation or are otherwise difficult to effectively parallelize. For applications with any significant level of random memory access, performance is primarily limited by bandwidth to main memory and between nodes. As processors become faster, unless memory and interconnects advance at the same rate, the processors must stall waiting for data for computation. While it is possible to achieve greater than 10% of peak (and this has been done), it typically requires "heroic" programming efforts.

Overall, memory system development for supercomputers has suffered from market economics. The demand for faster processors is broad, spurring tremendous improvements in microprocessor speeds. However, a relatively smaller percentage of the total market requires memory systems and interconnects with the potential to scale to processor counts greater than 1000 (or even 100). Unfortunately, a significant number of critical national security applications fall into this demanding category that is not well served by the commercial HPC marketplace today.

The prognosis for the clustered SMP class of machines is that current trends continue. Microprocessor speed and memory densities will continue to track Moore's law (doubling every 18 months) for the next decade; single address space memory bandwidth will continue to scale with frequency, but at a fraction of the rate required for balanced high performance computing. Memory latency will continue to increase relative to processor speed, further reducing processor utilization for codes that are not cache friendly. The design cost for microprocessors will continue to increase, further reducing the ability of commercial vendors to adapt their designs to the needs of the HPC market. The largest available single address space SMPs from most HPC vendors will remain stable at the current 64-120 processors sharing memory, largely because of lack of commercial demand for larger shared memory configurations, the difficulties of scaling the Operating Systems, and the lack of fault isolation in large shared memory machines[23]. Hence the performance for loosely coupled applications will continue to improve dramatically but little help is seen for large, tightly coupled, communications or random memory-intensive applications. The reliability of the SMPs themselves will become very good, but not nearly good enough for very large capability systems with thousands to tens of thousands of nodes, if programmed with existing models. Improvements in software recovery will be required to compensate for periodic component failure. System manageability will improve for moderately scaled clusters but remain challenging for the largest scale systems.

The type-C approach to building supercomputers invests more of its design effort and dollars into bandwidth between processors and to global memory. Consequently, relatively less leverage of the broader IT market is possible. This class of machines can typically achieve greater than 50% of peak performance on large, memory intensive applications; indeed for some memory

[23] Only one vendor is pursuing larger shared memory configurations, and this is in cooperation with NASA.

and communications-bound applications, these are the only platforms that can execute the task. Further, programmers can quickly achieve a substantial fraction of the ultimate potential performance of an application on this type of machine without heroic optimization efforts. Systems in this category are made today by only two vendors, one of which is foreign. The economic model that would support existence of this type of high-end machine relies on the ability of these products to "scale down" (i.e., the market competitiveness of smaller versions of the very high-end systems purchased for the largest national security applications). This is in contrast to the "scale up" approach of type-T architectures. The alternative economic model for type-C systems would have to involve the national security community purchasing sufficient quantities of such machines to contribute to an adequate revenue flow for the vendors. The consensus prognosis for the type-C class (and hybrids) is that no such machines will be produced absent government sharing in the non-recurring engineering costs, as well as the government expanding its role as a major customer. A balanced world-wide economic model for type-C class machines requires a growth in demand from the totality of increasingly complex computational science challenges of the scientific, social and governmental sectors, a reasonable but not measurable (at this point), prospect.

SYSTEMS SOFTWARE AND PROGRAMMABILITY

A primary finding of this working group is that high-end computers are too hard to use. The national security community will not be able to realize the full potential benefit of HPC unless progress is made in reducing the human resources cost to building and debugging large programs, tuning programs for performance, setting up applications, administrating the machines, and moving applications between machines of different types.

The baseline outlook for programmability is not healthy. There are virtually no forces driving improvement in software productivity for HPC systems. The dominant current models, including message passing (MPI) for clusters, and OpenMP for SMPs, will mature but become increasingly difficult to use as systems and applications scale. Key difficulties are the need to integrate two different models in one application and the need to reflect machine characteristics such as number of nodes, number of processors per node, memory size, and memory hierarchy in algorithm design. Distributed Shared Memory programming models like UPC and Co-array Fortran, while not the ultimate answer, are a step in the right direction. However, they would need significant government push to break through the chicken-and-egg problem of lack of significant demand for languages that are not broadly supported and lack of broad support for languages that are not widely demanded. This push must come in the form of substantial funding for tools and for the porting of multiple showcase applications into forms that demonstrate the advantages of these models and tools. Note that the time frame for development of sophisticated new tool sets to mirror new programming models rivals that of a development cycle for a new machine. In addition, changes are needed to the parallel programming models to tolerate the dynamic failure of one of the hardware building blocks of the system. However, even if significant progress is made with these models, parallel programming will continue to be more difficult and less well supported than sequential programming.

Cluster computers typically run one commercial operating system image on each node. A variety of cluster services, such as parallel schedulers or parallel file systems, are built atop the distributed operating systems using standard networking protocols. This "cobbled together" approach enables supercomputers to leverage the large investments in commercial operating systems, but results in poor system integration. This loosely coupled system model is a poor match to tightly coupled applications and a further deterrent to performance.

A third programming model popular today is that of the "computing grid" – using multiple nodes on the web (each of which may be large scale machines as discussed above) on the same application. Some success has been claimed for this approach (such as the SETI@home effort, which yielded arguably the first application to pass a petaflop). However, the bandwidth and latency between nodes is orders of magnitude worse than either of the previous system types, and is thus limited to problems having high compute to data set size. These are essentially those problems that can be broken down into a large number of completely independent pieces that can run without mutual interaction, a very small subset of applications. Significant problems also exist in managing such hugely heterogeneous systems, with different resources, operating systems, and computing capability.

The working group did not establish a roadmap for software technologies. One reason for this is that progress on software technologies for HEC are less likely to result from focused efforts on specific point technologies, and more likely to emerge from large integrative projects and test beds: one cannot develop, in a meaningful way, software for high performance computing in absence of high performance computing platforms.

The outlook for Input/Output (I/O) is only a little better. The next decade should see the maturing of Storage Area Networks (SAN) and Network Attached Storage (NAS), as well as parallel file systems. These file systems will extend to Global File Systems. That is, data will be accessible via a file system interface, independent of its physical location. The advent of OC192c (10 Gb/s) and OC768c (40 Gb/s), coupled with 10 Gb/s LAN's (10 GigE) will make this feasible. The NSF funded Teragrid is an example of this type of system-wide architecture. The convergence of WAN/LAN/SAN is beginning to happen.

While these developments improve the storage environment, the bottleneck of making multiple memory-to-memory moves before even the simplest I/O operation can take place chews up huge amounts of precious memory system and interconnect bandwidth within the system.

COMPONENTS AND BASIC TECHNOLOGY

Industry is pursuing basic technology research and development that is on the critical path for the next generation of systems. Heavy investment in bulk CMOS processing has sustained the performance improvements that fuels Moore's law. The 2001 ITRS Roadmap projects feature size scaling to 22 nm by 2016.[24] This prediction has significant technical challenges and the accompanying risks. An aggressive technology program will be required to continue Moore's Law scaling of CMOS. Incremental changes in operating temperature, interconnect approaches, etc. may be used to supplement process development. An aggressive, commercially supported, effort is underway, and high end computing does not drive it.

High Performance Computing is the beneficiary of an enormous research investment by the broader information technology economy. Hence, where the technology needs of desktop and midrange computing align, HPC is a beneficiary. To stop there would be very shortsighted. There is (with the exception of quantum computing) an insignificant level of basic technologies research and development in support of HPC systems more than five years in the future. This lack of research and development, especially at universities, has produced a knowledge gap; the United States is not producing the next generation of scientists that will support research and development for basic technologies for the High End Computing industry going forward. In short, the research pipeline has drained and needs to be reprimed.

Table 1 shows a matrix of potential technologies available for future HPC systems (first 9 columns) and, in the last column, the design and

[24] See ITRS Roadmap at http://public.itrs.net.

test tools that will be required. The candidate technologies can be lumped into 4 different classes:

Potential in less than 10 years (white boxes). With proper support for research and development these technologies could be ready for real deployment in less than 10 years.

Promising potential in more than 10 years (blue boxes). With proper support for research and development these technologies may be ready for real deployment in 10 years or more.

Not suitable for HPC (red boxes). These technologies are not expected to meet the needs of future HPC Systems.

Research and development of these technologies is currently being adequately addressed by industry or these technologies have reached a level of maturity such that no large performance increases are expected (green boxes).

Baseline technologies include most implementations of Si CMOS for logic and memory, magnetic and optical storage, multi-chip modules, boards and modules, electrical coax and ribbon cables, pliable interconnects, and air cooling. For many of the components and technologies listed in Table 1, the US has leading research and development. This lead is diminishing by a combination of aggressive advances from foreign suppliers and joint ventures between the US and

System Components and Technologies

Logic	Memory	Inter-connects On–Chip	Inter-connects Chip–Chip	Inter-connects Component - Component	Switches	Storage	Packaging	Thermal Manage-ment	Design and Test
Si-CMOS	Si-CMOS	Optical - Free Space, Guided	Optical - Free Space, Guided	Optical– Free Space, Guided, MEMs	Optical	Magnetic	MCM	Liquid-Single Phase, Multi Phase	Physical Design Tools
Si/Ge-HBT	Ferroelectrics	Cu-Low k	Low k	Stacked	S.C.	Optical	Stacking	Air	Layout
SC-RSFQ	Magnetic	RF	RF	RF	Semiconductor	Optical Tape	Wafer Scale	Cryogenic	Design Test
Optical	Optical	Heterogeneous Interconnects	Stacked	Electrical Coax Flat Ribbon	RF	Spintronics	HIST		Components Test
III-V-HBT	III-V	Carbon Nanotubes	Wafer Scale	Pliable		Scanning Probe	Boards & Modules		
Spintronics	Spintronics	Nano-electronics					Assembly		
Nano-electronics	Nano-electronics						Integrated Interconnect Technology		
S.C./Si-CMOS	S.C./Si-CMOS						Pliable Interconnects		

Potential <10 years (white)	*Potential in 10+ years (blue)*	*Not Suitable for HPC (red)*	*Mainstream development is adequate (green)*

S.C	superconductor	MCM	Multi Chip Module
S.C.–RSFQ	superconducting rapid single flux quanta	HIST	Heterogeneous Integrated Semiconductor Technology
III-V	three-five compound semiconductor	Cu	copper
HBT	heterojunction bipolar transistor	Low k	low dielectric constant insulation

Table 4 – System Components and Technolgies

foreign companies. If domestic innovation is not maintained, there will be less differentiation between advanced baseline technologies found domestically and overseas. US advantages could arise from innovative applications of the most advanced baseline technologies (e.g. 3-D wafer integration of advanced Si-CMOS), or the substitution of a new technology in place of the old (e.g. optical interconnects and switches in place of their electrical counterparts). In other baseline technology areas the US no longer holds the lead (e.g. high density boards and modules, lead by Japan) and this will not change under the current environment. Due to lack of research funding, the US will not be able to perform the truly innovative work required to leapfrog the incremental development of conventional boards and modules.

Though the U.S. has companies (like Intel and IBM) that are at the leading edge in silicon process and device technology, excellent process technology is also available from foreign suppliers. Increasing numbers of U.S. chip suppliers rely on foreign foundries like TSMC and UMC (Taiwan) and Chartered (Singapore) for IC manufacturing. Many leading edge suppliers, such as field programmable gate array (FPGA) vendors, have adopted “fabless” business models, and are reliant on these offshore foundries. In addition, due to the growing cost of process development and facility capitalization costs for advanced CMOS processes, many integrated device manufacturers (such as LSI Logic, Motorola, National, and TI) are outsourcing production to these foreign foundries, and in many cases totally rely on these large foundries for their most advanced processes. U.S. supercomputer makers will most likely rely on these foreign foundries to manufacture specialty silicon chips for their supercomputers. In the short term, excellent process technology is readily available and reasonably priced, opening opportunities for new supercomputer ASICs. In the long term, there is always the risk that the U.S. could find itself dependent on foreign suppliers who may not provide us with their best technologies. Currently, a relatively level playing field exists in process technology because both the U.S. and foreign chipmakers are dependent on the availability of leading edge process equipment to advance their process technology, and fiercely competitive suppliers in the U.S., Europe, and Japan sell this equipment. As a result, any country can “buy their way in” to leading edge production (but not necessarily to profitability).

Given the global availability of leading edge processes, supercomputer makers should adopt several strategies to gain a competitive advantage: First, they can design clever architectures and circuits to exploit those processes. Second, through the use of better design tools and talent, they can be first to market with improved designs. Third, they can invest some effort in intellectual property protection schemes to insure that their designs cannot be easily copied by a foreign supplier who gains access to their mask designs. Fourth, the U.S. can make investments in niche process technologies that can be used in conjunction with leading processes to improve their utility for supercomputing applications. For example, techniques to improve heat removal, shorten wire paths, increase circuit packaging density, reduce power, or operate at cryogenic temperatures are possible technologies which might be important extensions to silicon CMOS for supercomputer applications. Finally, the U.S. can continue to perform research in process technologies and CMOS devices to maintain on-shore expertise and intellectual property, insuring that the U.S. will continue to be a strong player in process technology and equipment, design tools, device and circuit design, and manufacturing. Despite the large R&D investments made by the semiconductor industry, well-targeted government funding directed at longer range research has dramatically affected the roadmap in such areas as advanced lithography.

The system–level performance metrics that the candidate technologies in Table 4 (page 41) must be judged against include:

High bandwidth/low latency interconnects and switches at all length and complexity scales in the system

High integration of previously disparate technologies (dense conventional logic, reprogrammable logic, dense DRAM memories, and non digital interface technologies such as optics) onto a single die

High clock speeds

Low energy operation

Reliability against both physical catastrophic failures and soft errors

Sustained high I/O rates at all levels of the system (both on-off chip and to peripherals)

Dynamic reconfiguration

Scalability of architecture

The most important performance metric identified by the SAPCWG as well as by the other working groups was the requirement for high bandwidth/low latency interconnects and switches, between logic, memory, and storage on both localized and distributed levels. Interconnects and switching between logic and memory have the highest performance demands. Distance and complexity make the interconnection and switching problem harder and more critical as HPC systems become increasingly larger. Commercial interconnect and switching technology is far behind Moore's Law.

The recommendations of the SAPCWG are that the majority of the technologies that fall into the 'feasible in less than 10 years' and 'promising in 10+ years' categories should receive sustained support for research and development. The critical *absence* of technological solutions in these categories has been recognized by the electronics industry and is clearly spelled out in their most recent International Technology Roadmap for Semiconductors. Special emphasis should be paid to interconnects, switching, packaging, thermal management and design & test tools. The reliability (especially from soft errors) of very advanced silicon device technology should be addressed. Non-silicon memory and nonmagnetic storage technologies should be explored.

QUANTUM COMPUTING

Quantum computers, based on controlled interactions of individual quantum states, are envisioned to provide a completely orthogonal method for solving a select set of problems which can currently only be considered with HPC systems. The promise of quantum computation is a drastic reduction in hardware required to solve certain problems that scale exponentially.

In considering the scope and timeframe for an aggressive HEC R&D program, the effect of quantum computing must be accounted for. In order to do this, several significant facts must be noted:

Only a select set of algorithms has been developed for quantum computation. At this point in time, it is expected that quantum computers will not displace either of the two types of capability computer systems: clustered symmetric multi-processors (SMPs) or clusters of scalable vector processors. Instead, the most likely outcome is the creation of a third type of HPC co-processor system that may be the system of choice for a small set of problems.

Barring an unforeseen breakthrough, a quantum computer of the scale necessary to attack problems of interest for national defense is not expected to be achievable for more than 20 years. While quantum gates have been realized, and quantum algorithms have been run, the components used for these demonstrations have not been shown to be scalable to the thousands of quantum gates necessary for real problems. One way of illustrating this gap between early demonstration and real systems is to note that current and foreseeable classical computational power is sufficient to simulate small (i.e., few quantum gates) quantum computers. It is entirely possible, for example, to simulate the operation of the largest current quantum computer—-the LANL group's 7-qubit NMR system—-on a desktop PC. To provide a computational advantage, any proposed realization must at least scale to a size beyond the simulation capability of any classical computer. Scalable approaches to quantum computing (one promising class being solid state semiconducting and superconducting circuits) are just beginning to be explored and single gates have not yet been achieved.

Algorithms for quantum computation may require a great deal of data preparation, quantum gate preparation, and post-quantum-operation analysis. The paradigm of an HPC system as an adjunct to a quantum computer or (more likely) a quantum computer as an adjunct to an HPC system is the most likely quantum computing paradigm. In any event, it is unlikely that quantum computers will replace any classical HPC systems.

Actually computing with quantum systems is, today, not well understood. There is no "quantum computer architecture" on which different algorithms could be run. There has been a lack of consideration of how one would couple quantum computers to conventional ones, allowing new data sets representing new problems to be inserted, and results read out for post analysis (which for several of the key quantum algorithms still requires significant "conventional" computing). Quantum computing is not well enough understood at this point to influence HEC R&D and its research should be considered outside the scope of the program under consideration here.

Technical Characteristics of an HEC R&D Program

In outlining the technical characteristics for an HEC R&D Program, the System Architecture, Programmability and Components Working Group wishes to tie together each of the major technical activities to ensure that each completely supports the broad objectives of such a program; i.e., the continued development and market insertion of the most capable high end computers that industry and government can collectively create. These technical activities include (1) intensively working with vendors to ensure that current and proposed offerings have the greatest capabilities that existing technologies will support, (2) enticing vendors into significantly different systems than they currently market, (3) investigating entirely new architectures and systems, developing those most promising and inserting them into the marketplace, (4) aggressively pursuing both hardware and software technologies, critical to the future of high end computing, in a comprehensive long term research program. The goal is an integrated approach to enabling U.S. leadership in this critical national security field.

The SAPCWG proposes the following major foci for the program:

Long-term basic and applied research in the key enabling technologies for HEC, both software and hardware. The research would occur primarily in U.S. universities, but include government laboratories and U.S. industry as appropriate.

Advanced Development of components and subsystems. This foci takes successful applied technology research and aggressively pushes through the next steps of development towards viability in future high end computers. Both software and hardware developments are critical and must be integrated.

Engineering and prototype development. As its core activity this program would partner with industry to develop new high performance computing systems, up through the production of serial number 1 of a new model. These projects would be competitively selected and may involve follow-ons to current vendor products that are feasible and required but would not be produced absent support from the national security community. They would also involve completely new systems that are pursued from concept through design and development and prototyping phases within the program.

TIME TO SOLUTION

Prior to further description of an Integrated High End Computing Program, an overriding concept is introduced for formulating the entire technical framework for the program: "Time to Solution".

Historically, high end computing systems have been acquired from the perspective of either reducing the execution time or providing memory resources that could not be provided in a lesser machine. However, reducing the execution time is only one component of "time to solution". The other component is "programmability," the time required to code the problem onto the computing platform, offload the results, and analyze them. Time to solution is often dominated not by system execution speed but by the speed of developing, modifying, and validating software. In experimental environments with rapidly changing computational requirements, the productivity of software development is as important as achievable computational speed.

In order to meet the ever-increasing demands for computing horsepower, we have acquired systems consisting of hundreds to thousands of processors; the NNSA/ASCI computing platforms are examples of such systems. In scaling such systems to meet future required levels of performance some are likely to contain hundreds of thousands of processors in an attempt at meeting computational demands; IBM's Blue Gene effort is an example of systems exploring this space, albeit for a specific class of problems. These systems are, however, only addressing the execution time component of reducing time to solution. Missing is the typically heroic efforts required to program these increasingly complex systems. Here "programming" represents all of the human-centric efforts required to: code, setup the problem, debug, analyze and recode. Indeed, timescales on the order of months or even years are required to program the highly complex codes required to address national security problems. The net result is that the potential contribution of high performance computing to national security as well as to such diverse segments of the economy as automotive design, medical and drug research, energy exploration, and biology, to name but a few, is not fully achieved.

Clearly, what is needed is an effort that addresses both aspects of time to solution, a balanced effort to facilitate the programming process as well as speed execution time. These must be addressed in concert; designing and developing the execution platforms capable of meeting the processing demands of the national security sector supported by a programming environment that sufficiently abstracts those computing resources to enable a significantly reduced programming effort. Any hardware initiatives must be accompanied by a co-evolution of software initiatives, as each depends upon the other. These challenges are illustrated in the figure 3 (next page).

Referring to Figure 3, the vertical axis represents the human-centric efforts covering all aspects of representing a problem in the computing resource. The horizontal axis represents the execution time on the computing resource. The bow shaped curves are lines of constant time-to-solution. Different applications may have different programming/execution splits. The further a time-to-solution is from the origin, the greater the time required to arrive at a solution. The goal is to drive DoD related applications closer to the origin, i.e., to reduce total time to solution.

Time to solution is multi-dimensional. As a result, any efforts at reducing the time to solution must address all of its aspects. What is required is an approach that provides efficient utilization of the computing resources (processors, memory, storage, communication) in addition to providing a comprehensive programming environment (new

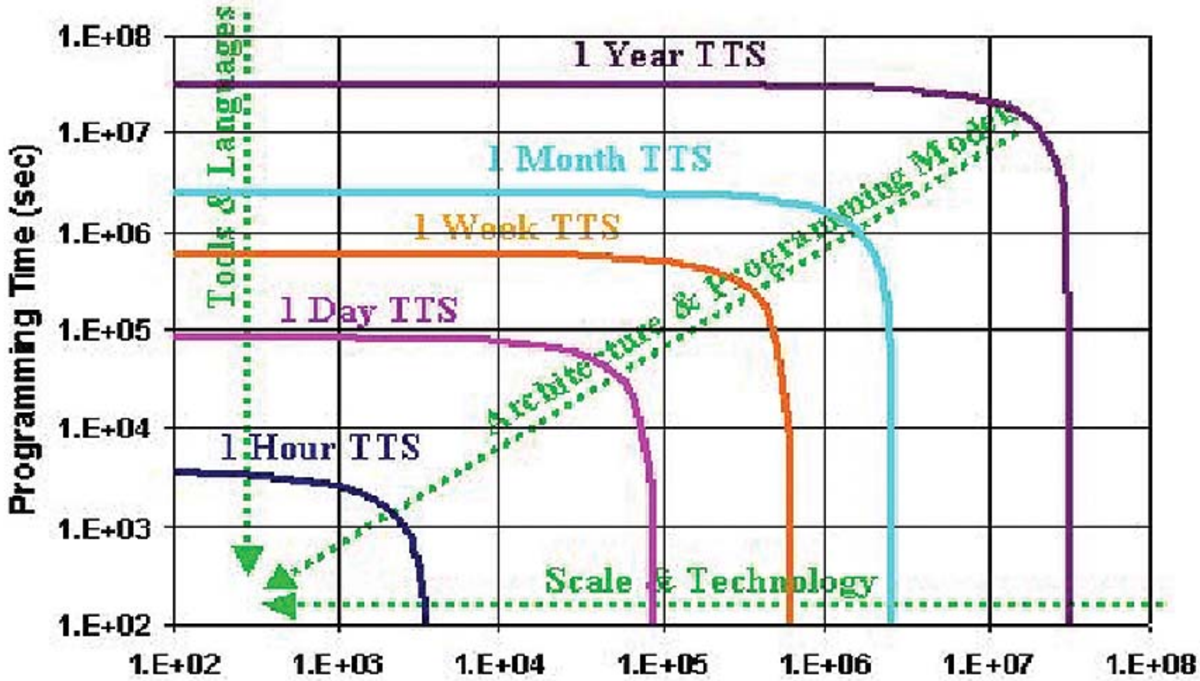


Figure 3 – Time to solution = (Human) programming time + (Hardware) execution time.

programming models, new language/compilers/ debuggers, middleware, libraries, etc.) to address the programmability of these complex systems. It is important to note that innovative system architectures may not only improve computation time (the horizontal axis), but also make machines easier to program, i.e. lower programming time (the vertical axis).

A key to improving programmer productivity is raising the level of discourse about programs. Some concepts do exist which address this goal. Higher levels of abstraction in program construction are an essential component. Manipulating entire objects (arrays, trees, etc.) with single commands, as in languages such as APL, merits further investigation. Self-describing tagged data structures would allow debuggers and other introspective tools to operate and manipulate programs and data in a flexible way at run time. Erasing the boundary between compile- and run-time allows incremental compilation, recompilation, and run-time optimization for performance by dynamically moving and restructuring code, data, or both. Simple language features such as physical units and dimensional analysis would provide tools that are (embarrassingly) missing in current computer languages. Strong protection of data objects through capability addressing would eliminate the vast majority of buffer overflow security attacks, and provide good "fences" between objects, localizing failures that might otherwise not be contained. Another key idea is the notion of transactional execution. While currently employed in the reorder and commit stages of modern, out-of-order processors, little support exists at higher levels of software design. Transactional execution could be extremely important in the graceful handling of the inevitable hardware and software faults in large systems. Finally, it is essential that the key problems facing today's computers, memory latency and bandwidth, be recognized up front, and made explicit parts of the discourse. Alternative ways of tying computation to objects that avoid the bottlenecks need to be more fully explored.

The enabler for systems providing these capabilities is the abundance of processor resources available (relative to the scarce resources of interprocessor and memory bandwidth and programmer time). With transistors no longer the critical resource, the balanced system of the future can devote a much larger fraction of its resources to raising the productivity of its programmers.

Another key contributor to programmer productivity is the design of parallel hardware with more robust performance characteristics. Higher interprocessor bandwidth and lower interprocessor latency, and a flatter memory hierarchy not only enable high performance for those applications that need to communicate; they also avoid the heroic efforts needed to achieve acceptable performance, even for those applications that can be suitably decomposed.

Thus, a critical issue across all major segments of the Program will be to challenge all initiatives to describe improvements in time-to-solution, be it in leveraging up vendors' current products, improved or radically new architectures and systems, and technology enabling R&D. In particular, programming environments and new programming methodologies will be a critical element of the long-term research component of the program, as well as advanced development and prototyping efforts.

FOCUSED DEVELOPMENT PROJECTS

There is no better way to drive the HEC R&D Program than to aim it ultimately at delivering systems at a target date. It is therefore the finding of the SAPCWG that the national security community needs an on-going, multi-disciplinary effort to incubate innovative architecture concepts supported by appropriate technology development efforts culminating in constructing prototype high end supercomputers that match user community needs and are commercially viable. The proposed model is a spiral consisting of a process that funds a number of completely new and innovative architectural concepts that demand leading edge, long lead time technologies. The more promising new systems (less than 10 years out) would then proceed to a proof of concept or test bed phase in which components and small systems are developed in parallel. This is likely to be primarily a university-based effort in its early phases.

At this point, a number of candidate architectures with a market time horizon of <5 years compete to proceed to prototype development. Test beds and prototypes are critical to avoid the phenomenon of designing to the lowest common denominator; if new systems need to be able to run critical applications, all technical risks must be mitigated before production begins. Further, today's supercomputers are too complex to allow a model of testing new ideas in the market place, and innovative ideas cannot be tested without a reasonable scale prototype. Test beds may be smaller and slower than the ultimate system they model, but they are large enough to test scaling and system software. They are heavily instrumented with perhaps as much money spent on instrumentation as on the installation itself. Projects in this phase must focus on software and program development environments as well as hardware systems. Consortia of research labs, industry, and academia accomplish the prototype phase. It must be emphasized that test beds are for experimentation and are not operational platforms.

A key component of the test bed efforts is risk tolerance and management. Projects that depend on too many high risk components are likely to fail. At the other end of the spectrum, projects with little risk are not research and are more properly built as products by vendors. This effort must artfully seek the middle ground in which some risk of failure is tolerated (as it must be in any research effort) but projects are not doomed to failure by dependencies on multiple high-risk technologies.

Our concept for development requires that architects, systems designers, software developers and certain technology initiators have available large-scale simulation capability. We have had extensive experience in development of new systems and have seen, over and over, delays in development of systems software until a reasonably sized prototype system is available. Most individual system developments cannot afford the cost or time delays inherent in creating system-specific reasonably sized simulators. A centralized large-scale general-purpose simulator will save overall systems development expense and compress system development schedules. The simulator will be housed and supported at a HEC center-of-excellence, capable of other centralized research, configuration management and support functions.

In the final phase, operational platforms based on successful prototypes, are built by industrial contractors, and procured and deployed by the mission agencies. It is essential to the health of the high-end computing environment in the United States that Defense mission agencies have the intent to purchase instances of these final phase machines. However, it is recognized that such agencies have missions to perform and will always need to make procurement decisions based on accomplishing that mission. In a healthy steady state, these final phase machines will be the next generation products of U.S. HPC vendors and will be the best machines for accomplishing the top end high performance computing requirements of the Defense mission agencies. We confidently expect that the federal user community will carefully track the process outlined here and orient their procurements to these systems.

Good models for collaboration with high end vendors to bring advanced machines to fruition exist. The NNSA/ASCI collaboration with IBM on the Blue Gene/L machine and DDR&E and NSA's collaborative R&D effort with Cray, Inc. on the SV2 are examples that culminate in mission production machines. The NASA Ames/SGI cooperative development of the 1024 processor SMP node and several NNSA/ASCI Path Forward projects are smaller scale examples of mission-driven technology insertion efforts. Another excellent model for the interaction with the high end industry is DARPA's High Productivity Computing Systems Program (HPCS). This program challenges vendors, usually with broader academic and multiple-firm participation, to develop systems concepts a stage beyond current market systems. The concept proposals are competitively selected, proceed through a funded design stage, and may be further down selected to one or two systems that will be funded through completion. During this process, there will be intensive discussions with vendors to insure adoption into vendor product lines. These existing program models should form a basis for designing the Integrated HEC program.

LONG TERM RESEARCH AND APPLIED TECHNOLOGY DEVELOPMENT

One of the many weaknesses of the high end computing industry is the absence of significant relationships with academic institutions capable of feeding new and innovative research results into the chain of experimentation, development, prototyping and engineering of new systems. Centers of excellence affiliated with major educational institutions sometimes bridge this gap but results are spotty. Major institutions such as the National Science Foundation fund some basic research but are more focused on applied computational science rather than research in high end computing technologies and systems. In any event, the NSF and similar institutions are all but forced to stay well away from commercialization of research results.

The HEC R&D Program must pick its way through this maze. We must take full advantage of work being funded by other sources, we must develop and maintain strong linkages to academic centers of excellence in critical fields, and we must maintain, indeed improve, current partnerships with NSF, DOE Office of Science, and other federal non-national security agencies in high end computing areas of common interest.

The long lead–time applied technology research program should incorporate research programs in at least the following seven fields:

System Architectures

Memory Subsystems

Parallel Languages and Programming Environments

Packaging/Power/Thermal Management

Interconnects and Switches

Storage and I/O

Novel Computation Technologies.

At present, very few HEC technology development efforts are underway under current R&D programs. The Program envisioned in this paper anticipates creation of a Long Term Research Component to support the program in all its phases. It is neither appropriate nor desirable that the HEC R&D Program duplicates any technology effort currently underway in industry. Rather, the Program must concentrate on those technologies that are not being pursued, yet are critical to our work with vendors on existing systems, vendor future systems and new and innovative architectures.

Table 4 summarized hardware system components and technologies of importance to high end computing; basic software research is equally critical. Technical issues that will dominate the HEC R&D Program technology efforts will include:

Power. Energy consumed per operation is a critical metric as high-end computing becomes increasingly power constrained. For example, the Japanese Earth Simulator, a processor, already consumes 5MW to achieve 40 Teraflops. There are a number of tradeoffs that can be made to reduce the switching energy for a CMOS gate by >100X from today's values. Total power consumption is a limiting factor affecting large systems.

Developing terabyte memories with a reasonable cost per bit and snuggled close enough to the processors to keep latencies low and bandwidths high. Advanced packaging technologies for HEC require special attention by the HEC research community.

Component technologies significantly lagging others. The key gap that must be addressed involves memory subsystems, particularly on/off chip latencies. Processor-in-Memory (PIM), 3-dimensional configurations, cryogenic cooling, and innovative arrays with improved chip-chip interconnections are but a few of the progressive technologies that show good experimental results. Processor-memory mismatches and memory bandwidth issues are not significant drivers for the general computing industry; yet these issues cripple performance growth in most high end computing architectures and it appears inevitable that these issues will be high on the HEC R&D technology agenda.

Software and hardware component technologies that are likely to be unique to high end computing, at least initially. These technologies are likely to wallow in research land for want of compelling commercial incentives. Examples include several technologies mentioned above but also advanced optical interconnection technologies, complex networking topologies and associated optical switching technologies, and a host of software and programming technologies that have little or no appeal to ISV's.

Key to a technology development program will be partnerships with component technology companies and incentives to vendors to follow developments closely, particularly when they are critical elements of new architectures intended to migrate into commercial production.

We will forge close technical relationships between and among our national security research laboratories, academic institutions and industrial centers-of-excellence. The research agenda will be visible, focused on the long-term needs of the national security high end computing capabilities, yet sufficiently insulated from pressures to permit full exercise of the intellectual resources of the research organizations.

SAPCWG Conclusions

The SAPCWG working group has analyzed the requirements for future high end computing systems identified by the OUWG. We concur with the OUWG's findings that the computational characteristics that limit current and future high performance computing systems are: memory subsystem performance, programming environments, CPU performance[24], and I/O subsystem performance. We agree that both type-T and type-C systems play critical roles in current national security capabilities and that the potential for enhanced security in the future will rely on increasing HEC capabilities and supply. We find that the U.S. is likely to continue to lead the world in type-T systems, but that research and development is still required to bring the memory systems and especially the programming environments up to the level required by national security mission agencies. We find that a domestic source of type-C systems is unlikely to be maintained by market forces alone. We further find that the potential for new computing technologies and platforms to come to market is not high without substantially increased support for basic research, technology incubation, and prototype development.

We have described characteristics of an integrated program that addresses long term research, technology and component incubation, and the engineering and prototype development of whole supercomputers. The vision supports close ties between national security mission agencies, the U.S. industry that produces computers and components, and the academic and broader research community. We are confident that, with such a program, a healthy U.S. supercomputer industry and research base will enhance our national security.

[24] The SAPCWG would generalize CPU performance to include the whole processing architecture.

APPENDIX A - STUDY PARTICIPANTS

Study Leadership

Charles Holland, Department of Defense

William Reed, National Nuclear Security Administration

Dan Wolf, National Security Agency, Chairman

Senior Advisor

George Cotter, National Security Agency

Technical Leaders

Robert Graybill, Co-Chair PWG
Cray Henry, Co-Chair OUWG
Gary Hughes, Co-Chair PWG
Ray Miller, Co-Chair OUWG
Jose Munoz, Co-Chair SAPCWG
Thomas Page, Jr. Co-Chair SAPCWG
Edward Pate, Co-Chair OUWG

Academia

Keren Bergman, Columbia
Kenneth Kennedy, Jr., Rice
Peter Kogge, Notre Dame
Thomas Knight, MIT
Robert Lucas, USC
Kevin Martin, Georgia Tech
Paul Messina, Cal Tech
Daniel Reed, U. of Illinois UC/NCSA
Thomas Sterling, Cal Tech
David Shaver, MIT/LL
Marc Snir, U. of Illinois UC

Government

Robert Alexander
Fernand Bedard
William Burnett
Paul Boudreaux
Mike Clancy
Robert Ciotti
Candace Culhane
Hank Dardy
Larry Davis
Robert Guerriero
John Grosh
Robert Herklotz
Alan Hunsberger
Frederick Johnson
Roman Kaluzniacki
Robert Karkoska
Pamela Krause
Robert Meisner
Michael Merrill
Baron Mills
John Parmentola
John Placanica
Ray Price
Tom Richardson
Hans Ruppel
Lauren Smith
Frank Thames
Phil Webster
Nancy Welker
Gary Zelanko
John Ziebarth

Industry

Lory Arghavan, Lockheed-Martin
Greg Astfalk, HP
William Carlson, IDA/CCS
Raymond Cosner, Boeing
David Goldschmidt, IDA/CCR
Charles Hayes, HCS
Richard Kauffman, Compaq, Inc
Alexander Larzelere, II, SAIC
John McCalpin, IBM
Allan Murray, IDA/CCS
Alan Piszcz, Mitre
Steven Reinhardt, SGI
Joshua Simons, Sun
Burton Smith, Cray
John Spargo TRW
Francis Sullivan, IDA/CCS
Joseph Swartz, Lockheed-Martin
Steve Wallach, Chiaro
Barbara Yoon

National Laboratories

David Bailey, Lawrence Berkeley
William Camp, Sandia
Lawrence Cox, Los Alamos
Robert Leland, Sandia
David Nowak, Lawrence Livermore
Doug Post, Los Alamos
James Rathkopf, Lawrence Livermore
Bob Thomas, Sandia
Joseph Thompson, Los Alamos
James Tomkins, Sandia
Mark Seager Lawrence Livermore
Sheila Vaidya, Lawrence Livermore
Andrew White, Los Alamos

APPENDIX B - STUDY CHARTER

Comprehensive, Long-Range, Integrated High End Supercomputing Program

PROPOSED STUDY ORGANIZATION

In the House Appropriations Committee Report to accompany the FY02 Defense Appropriation Bill, Congress tasked the Secretary of Defense with producing, by July 1, 2002, a development and acquisition plan for a comprehensive long-range integrated high-end supercomputing program. Congress further directed that the study will be lead by the National Security Agency and involve the cooperation of DARPA, DoD's HPC Modernization Program Office, NIMA, NRO, and NNSA/ASCI.

The study will result in a plan for a high-performance computing program, with options identified and ranked, in sufficient detail to support definition and implementation of the program following approval of the plan. The report describing the plan will be unclassified, with classified appendices as necessary.

A Senior Group will be formed to provide overall guidance and oversight to the effort and three working groups will also be formed to work cooperatively to complete the effort. The working groups include: an Operational Users Working Group (OUWG); a Systems, Architecture, Programming, and Components Working Group (SAPCWG); and a Planning Working Group (PWG).

The core membership of these working groups will be drawn from the organizations named above and with additional representation invited to participate from the high performance computing community to include industrial, academic and government; researchers, hardware and software developers and users of critical applications.

No advice, recommendations, or opinions will be requested or accepted from non-Federal employees, only information and statements of fact.

This remainder of this document summarizes the organization of the study in order to complete the task within the timeframe.

Senior Group

Membership: The Senior Group is comprised of General Officer-Senior Executive Service level members from NSA, DUSD S&T and NNSA/ASCI and will be chaired by NSA.

Responsibilities: The Senior Group has overall responsibility both for conducting the study and for presenting its results for coordination and approval. Specifically, the Senior Group:

- will oversee and provide continuing direction to the working groups; resolve or attempt to resolve all issues that arise, including issues with individual agencies at the agency level; approve any changes in membership, chairmanship, or responsibilities of the working groups; and ensure that only facts or information are solicited while any of the study working groups are meeting with representatives from more than one non-federal organization.
- will perform the final internal review of the plan, options, and ranking of options to be presented to decision authorities.
- is responsible for coordination and presentation of the results of the study.
- will coordinate at the agency level, present the results to the appropriate approval authorities (initially USD AT&L), and
- will perform any ancillary briefings that may be required.

STUDY PROCESS

- The Senior Group members will meet regularly during the study. Whenever possible, the Working Group chairs will participate in the meetings of the Senior Group. The Senior Group will make decisions by consensus as much as possible within the intent of the Congressional task and within the exigencies of the required completion date.
- To insure that OSD and Congress are provided with all relevant information, supporting analyses and the benefits of a diversity of opinion, the final study report will contain, as appendices, the full and complete work product, analysis and minority opinions from each working group.

COORDINATION AND APPROVAL PROCESS

- Each Senior Group member will initiate coordination within his agency to establish his agency's position and comments on the plan. There will be a single coordination phase at the agency/departmental level involving meetings of the agency principals as necessary.
- After this coordination phase, the Senior Group will present the DIRNSA signed report to USD AT&L for approval. If there are unresolved departmental coordination issues at that point, it is expected that they will be addressed at the Secretary level.
- Following review and approval by USD AT&L and depending on funding issues, it may be necessary for the plan to go through the Defense Resources Board process.

Operational Users Working Group

Membership: The OUWG will include representation from the major national security community user organizations and will be co-chaired by NSA, DoD Modernization Office and NNSA/ASCI. Members will be drawn from NSA, NNSA/ASCI, NASA Ames, ASD C3I, DoD Modernization Office, Naval Oceanographic (Fleet Numerical), NIMA, NRO, and the Military Departments high-end computing laboratories. It will also include representatives from Defense firms in cases where national security high performance computing tasks have been wholly outsourced to those firms (such as SAIC, Boeing, and Lockheed). The industry representatives may be excused from certain meetings or government only deliberations, as legal, regulatory or proprietary circumstances require.

Responsibilities: The OUWG will outline the program development (ease of use) and computational (capability and capacity) demands for critical HEC system characteristics for at least one challenge application in important national security sectors (e.g., cryptology, information processing, imagery, operational weather forecasting, combat aircraft modeling and simulation, nuclear stockpile maintenance, ballistic missile defense, battlefield management, major weapons design programs).

The OUWG will focus on the computational requirements and will seek to avoid taking any position on specific architectures, candidate technologies, or particular vendors.

Specifically, the OUWG will project, individually and collectively, the high performance computing system characteristics that will be essential to meet their most demanding challenge applications in the 2008-2010 timeframe. A *challenge* application is one that is critical to performing the mission of the agency, maximally demanding, and typical of that agency's demanding applications. This will serve the following purposes:

- To document the challenge applications for the plan reviewers and approvers.
- It will provide input to the other groups as necessary to support their deliberations.
- The documented challenge applications will serve as benchmarks for periodic review of the program objectives and directions.

STUDY PROCESS

- The OUWG, in consultation with the SAPCWG, will immediately identify an inventory or matrix of relevant system characteristics (e.g., ease of use, processor speed, bandwidth, memory size). This matrix will specify the system parameters whose values must be identified to determine what demands the challenge applications will place on future systems, architectures, component technologies, and materials science; in short, the program's technical objectives.
- Each member of the OUWG will then be responsible for identifying one or more challenge applications for their own agency, and for providing/estimating the system computational parameters for that application. The OUWG will collectively review these challenge applications to determine the range of system characteristics that are critically needed in the 2008-2010 timeframe.
- Throughout the process, the chairs of the OUWG will coordinate closely with the chairs of the SAPCWG and the PWG.

OUTPUT

- Technical data required by the SAPCWG and PWG, in the form of completed matrices of required system characteristics.
- Interim technical reports, analyses or working papers to support close collaboration and synchronization with the other groups.
- Weekly status reports to the Steering Group.
- A written report to be used as the *Requirements* section of the final report.
- Briefing materials to be used in creating the final briefing of the plan.

Systems, Architecture, Programming & Components Working Group

Membership: The SAPCWG will be co-chaired by NSA and NNSA/ASCI. Membership will include all agencies engaged in major HEC R&D activities, including NSA, DARPA, and NNSA/ASCI. Nationally known experts in HEC from industry and academia will also be included.

Responsibilities: The SAPCWG is responsible for identifying governmental, industrial, and academic solutions that are essential to address the computational requirements specified by the OUWG.

Specific tasks include:

- Document the baseline of current national security HEC R&D activities, including architecture, system, software and component efforts.
- Project industrial, academic and governmental research and developmental best efforts on critical technologies in support of high end supercomputing to the 2008-2010 timeframe.
- Identify critical gaps/shortfalls that occur when these best effort projections are considered relative to the challenge requirements identified by the OUWG.

STUDY PROCESS

- The SAPCWG will review existing documents and solicit information through data calls and briefings as necessary to complete the tasks enumerated above. It will solicit leading vendors of high performance computing systems to present briefings on their current and future capabilities and plans.

- Throughout the process, the co-chairs of the SAPCWG will coordinate closely with the chairs of the OUWG and the PWG.

OUTPUT

- Technical data required by the PWG in the form of community best efforts contributing to complying with the requirements of the Federal HEC community.
- Interim technical reports, analyses or working papers to support close collaboration and synchronization with the other groups.
- Weekly status reports to the Steering Group.
- A written report.
- Briefing materials to be used in creating the final briefing of the plan.

Planning Working Group

Membership: The PWG will be co-chaired by NSA and DARPA. Membership will include DARPA, DoD's HPC Modernization Program Office, NIMA, NRO, NNSA/ASCI, and NSA. The co-chairs of the PWG will be ex officio members of the OUWG and the SAPCWG.

It is expected that membership of this group will include key Government members from the OUWG and the SAPCWG who move into supporting the Planning Working Group as the analysis and deliberation phases within the other two groups are completed.

Responsibilities: The PWG is responsible for identifying the intersection between requirements and solutions identified by the other working groups. It will then highlight the shortfalls industry and academia will not address on their own, and clearly define those shortfall areas in which the USG can, and cannot, reasonably influence by a long-range integrated high-end supercomputing research and acquisition program.

The PWG is ultimately responsible for synthesizing the information, and analyses of the other two groups into a coherent program plan. The PWG, including Government members who transition from the OUWG or the SAPCWG, will be responsible for applying the results of the previous efforts to the task of compiling the study report and draft program plan. This includes:

- Incorporating the written reports from the other two groups as sections of the final report. The PWG will not evaluate or revise the positions taken the other two groups.
- Identify opportunities related to the gap analysis from the SAPCWG in which the USG can, or cannot, play a role.
- In coordination with key members, formulate a balanced long-term HEC R&D technical program proposal.
- Laying out the threat to national security interests from foreign competition, foreign dominance, and offshore dependencies. (Much of this material will be readily available to the PWG as the result of a separate Congressional question on this subject, which NSA is now researching.)
- Identifying the options, and the pros and cons for each program option, including:

  Program structure

  Program management

  Funding
- The PWG will identify but not resolve issues that arise in these three areas. It will recommend a ranking of the options, but it is not required to reach consensus on such a ranking; ranking recommendations will be reviewed and finalized by the Senior Group.

STUDY PROCESS

- The efforts of the PWG will be phased in as the study progresses. It is assumed that certain detailed studies and program and report framework tasks will begin at the outset of the effort and additional tasks will be added over time with the bulk of the work occurring near the end of the study period.

- The PWG will identify candidate options for program structure, management, and funding. It will obtain information on such options as currently instantiated in comparable joint research, development and acquisition programs to identify the pros and cons for each.

- The PWG will prepare the sections of the report for which it is responsible, drawing on existing information and reports where possible (e.g., the NSA report on foreign high performance computing systems that is now in preparation).

- The co-chairs of the PWG will coordinate closely with the chairs of the OUWG and the SAPCWG. The PWG must ensure that the planning options it identifies are relevant to the user requirements and the proposed technical program.

OUTPUT

- Interim technical reports, analyses or working papers to support close collaboration and synchronization with the other groups.
- Weekly status reports to the Steering Group.
- The proposed final plan report for review and approval of the Senior Group.
- The report will include as appendices the full and complete work product, analysis recommendations and minority opinions from each working group.
- The proposed final briefing, for review, modification and approval of the Senior Group.

Bibliography

There are a number of documents spanning the past several years that provide invaluable foundational material for this study. These documents will be made available at the outset of the study to all members.

- National Security High End Computing Integrated Process Team for National Security High Performance Computing (viewgraphs), 1996.
- Review of the National Security High End Computing Integrated Process Team Report on National Security High Performance Computing, MIT Lincoln Laboratory, 13 December 1996.
- President's Information Technology Advisory Committee (PITAC) Report on Information Technology Research: Investing in Our Future, Chapter 3, pp. 45-54, February 1999.
- Leading-edge Research and Development for Future Generations of High End Computing, HECC Coordinating Group, 1999.
- Leading-edge Research and Development for Future Generations of High End Computing, HECC Coordinating Group (executive summary for Bush Transition Team), November 2000.
- Report of the Defense Science Board Task Force on DoD Supercomputing Needs, 11 October 2000.
- NSA Response to the Defense Science Board Report, 22 December 2000.
- Survey and Analysis of the National Security High Performance Computing Architectural Requirements (viewgraphs), Richard Games, 26 April 2001.
- Survey and Analysis of the National Security High Performance Computing Architectural Requirements (report), Richard Games, 4 June 2001 (NOTE: Received from ODUSD S&T; not released for public distribution).
- HAC language in the Report to Accompany the FY02 Defense Appropriation Bill.
- The Role of High Performance Computing in Industry (Executive Summary), International Data Corporation, 1998.

APPENDIX C - DETAILED PROJECT DESCRIPTIONS

APPENDIX C – DETAILED PROJECT DESCRIPTIONS

CONTENTS

C-1 COMPREHENSIVE AEROSPACE VEHICLE DESIGN

Project Description

DoD war fighters critically depend upon air vehicles for carrying out their missions effectively and efficiently. Each of the three Services has large programs to develop future aerospace vehicles that will extend U.S. dominance in that arena into the coming decades. High performance computing is playing an increasingly vital role in current programs such as the F-22, V-22, and Joint Strike Fighter (JSF). It is of even greater importance as we look to future aerospace development programs involving hypersonic capabilities that will be much more difficult to test and for which more comprehensive physics must be added to current models in order to accurately simulate these flight regimes. The new hypersonic initiative will produce a vital baseline of technologies required to produce future dominant war fighting systems.

Design of aerospace vehicles is of necessity, a multi-disciplinary effort. Current capabilities to model the external airflow, propulsor performance, vehicle signature, and materials properties allow reasonable predictive results on today's HPC system when computed independently. There are a number of computational projects within the DoD laboratories and test centers that address these issues. However, the need to combine these independent modeling efforts into an interactive modeling capability in which all aspects of the design features are tightly coupled is one of the primary drivers for exponentially growing requirements for HPC capability in the future.. For example, the effect proposed changes have in the propulsor design on the vehicle signature can immediately be determined. Again, this integrated modeling capability will become of even more crucial importance to accurately model systems based upon new hypersonics technologies.

What is the real life problem?

The National Aerospace Initiative develops and demonstrates technologies for hypersonic systems. In the near term for DoD, hypersonic systems include hypersonic cruise missiles and in the future, hypersonic strike or reconnaissance aircraft and affordable on-demand access to space vehicles. Hypersonic cruise missiles would prosecute time critical targets, flying hundreds of nautical miles in minutes with cruise speeds of Mach 6 to 8. Hypersonic strike or reconnaissance aircraft would provide tactical commanders with near real-time reconnaissance information on demand. Cruise speeds would be between Mach 5 and 8. Hypersonic-powered access to space vehicles would enable true on-demand access to space to meet military operations tempos.

Hypersonic propulsion will enable a broad range of capabilities for war fighters. Significant advances in scramjet propulsion have occurred, but flight testing is required to clearly demonstrate the performance offered by these systems. Ground test facilities can operate at the appropriate Mach numbers for limited duration and do not have the capability to address the range of Mach numbers required to test transient engine operation. Computational techniques are required to assess scramjet engine operability from Mach 4 through Mach 8 in the near term and from Mach 10 through Mach 15 in the future. These efforts will enable efficient development and demonstration of scramjet engines and then combining the scramjet with rockets and or high-speed turbine engines capable of operating efficiently over the full range of Mach numbers.

One of the biggest challenges to the accurate prediction of external airflow is proper physics for handling massively separated flows. These predictions of massively separated flows over aircraft have been inaccurate and unreliable, causing aircraft designers to rely solely on expensive flight tests. Massively separated flows occur on aircraft maneuvering at high angles of attack and on bluff bodies. Current examples of massively separated flows include the V-22 in a descent, the F-22 in a high-g turn, and the F/A-18E at transonic speeds (abrupt wing stall). A hypersonic vehicle on re-entry is an important future example of a massively separated flow.

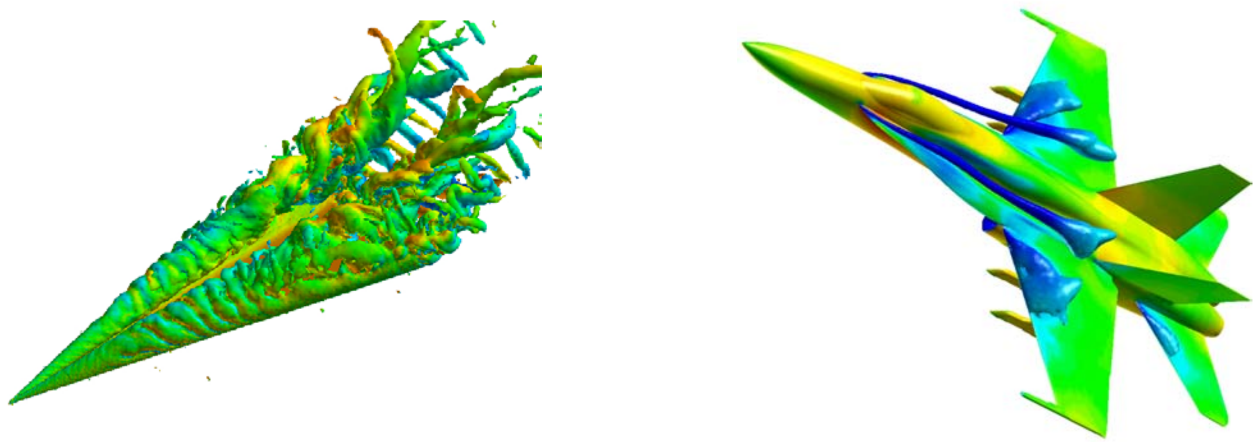


Figure 1: Vortex Breakdown on a delta wing, and the F-18C. This phenomenon causes aero-elastic bending and potential failure of the tails. The delta wing solution is using DES, showing capability to resolve unsteady flow features crucial to prediction aero-elastic flows. The F-18C prediction is using the previous state-of-the-art RANS methods, and is insufficient for aero-elastic predictions.

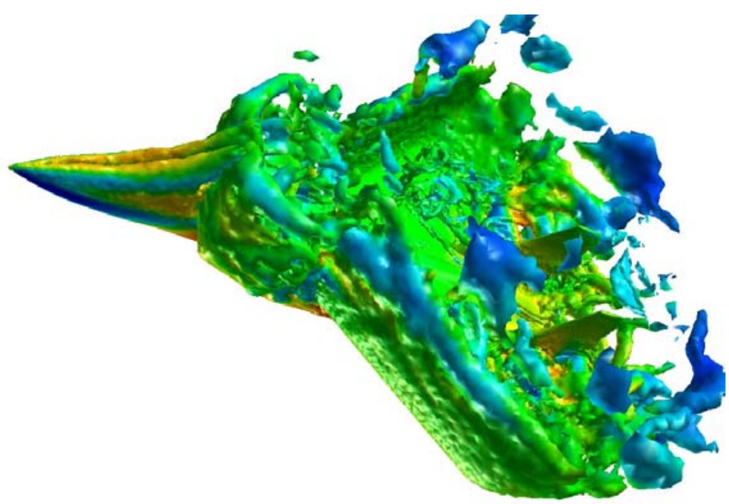


Figure 2: Unsteady flow over the F-15E at 65 degrees angle-of-attack. DES has matched flight test lift, drag, and pitching moment to within 5% on a 20 million cell grid. This gives DES the sufficient accuracy to predict spin, assuming a sufficient grid density.

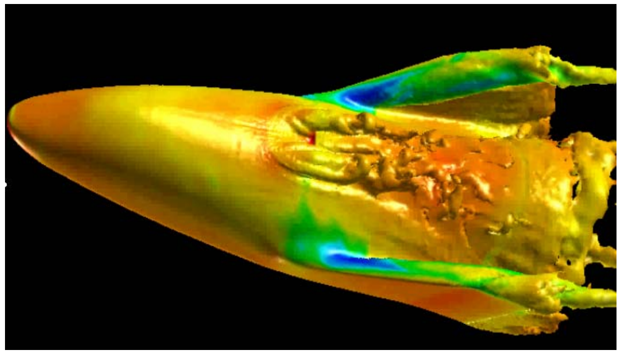


Figure 3: DES prediction of the flow over the X-38 hypersonic crew re-entry vehicle. DES is able to capture the unsteady flow features due to separation off the docking ring and afterbody.

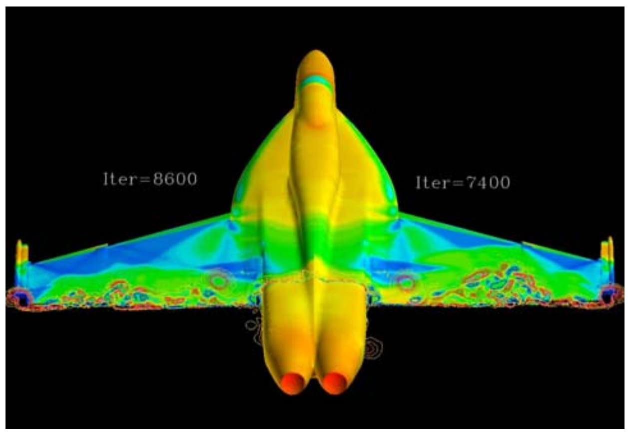


Figure 4: Prediction of unsteady shock oscillation on the F/A-18E using DES.

One current and future DoD aerospace system that has recently proved its worth in Afghanistan is the unmanned aerospace vehicle (UAV). More efficient and effective design of these valuable systems depends critically on the need to accurately model their nonlinear aerodynamics in high-altitude and transonic flight. This ability, in turn, depends on our ability to solve fluid-structure interactions in a truly interactive, interdisciplinary way.

Another major area of aircraft design for systems such as the JSF is to accurately model store carriage and separation. Each new aircraft-store configuration for current or future aerospace platforms must be certified, usually with costly and extensive flight tests. More accurate and efficient computational fluid dynamics (CFD) methods are beginning to allow certification with a more limited set of flight tests by providing the ability to target the tests toward key points in the flight envelope that are important to overall performance and safety. As the number of possible aircraft-store separation combinations increases, which is particularly important for joint platforms such as JSF, CFD modeling will clearly play a much larger role in the certification process.

For design of engines for current and future aerospace vehicles, thermal barrier coatings of turbines permit higher temperature operation than the single crystal metal superalloys could otherwise withstand. The increased combustion temperatures permitted by these thermal barrier coatings result in increased power and fuel efficiency of the engine. Likewise, these coatings protect the metal alloy from hot oxidation and corrosion. However, these thermal barrier ceramic coatings tend to crack and chip off the metal alloy after a series of engine operation cycles. This ruins the protective element of the coatings and can become dangerous if large fragments spall. The underlying causes of this spallation need to be investigated. In order to prevent or inhibit this failure, chemical modifications need to be made to the thermal barrier coatings. Computational materials science is playing a crucial role in providing this understanding that will allow more efficient and less-costly-to-maintain aerospace vehicle propulsion systems.

What is the National Security Impact of (solving and not solving) the problem?

Real world experiences in both Desert Storm and Afghanistan have shown that the U.S. needs improvements in targeting and weapons to destroy time critical targets. Improvements in targeting to significantly reduce the timeline from detection to weapons release are underway. The development of hypersonic weapons is critical to defeating these targets, reducing the timeline for the final link in the kill chain. Without hypersonic cruise missiles, the U.S. will be able to prosecute only a very limited portion of the total target set. Hypersonic strike and reconnaissance aircraft will combine high speed, high altitude, and moderate signature reductions to operate, with relative impunity, out of the range of enemy integrated air defenses. If such hypersonic systems are not developed, intelligence will rely on satellites, which are predictable and offer limited coverage, or tactical reconnaissance assets, which require additional air assets to ensure their survivability. High-speed strike systems, by virtue of the same survivability enhancements, will be able to penetrate enemy air defenses and strike targets at will. Failure to develop such systems will require tactical commanders to rely on conventional systems will additional air assets to ensure their survivability. Current access to space systems require excessive lead times to support planned military operations tempos in the 2025 time frame. Air breathing hypersonic propulsion will enable systems with true aircraft-like operations to support emerging military requirements. These systems will significantly reduce the cost per pound to place satellites in low earth orbit and improve reliability to become comparable to fighter aircraft.

Having the ability to predict massively separated flows around aerospace vehicles could substantially reduce acquisition cost and reduce the cost and risks of flight test. For example, the cost of flight testing the F/A-18E for abrupt wing stall could have been greatly reduced if this capability had existed several years ago. Correctly modeling this phenomenon will potentially accelerate development time for new hypersonic weapons and vehicles as a greater percentage of the flight envelope can be explored in a more timely fashion with high performance computing.

Accurate, integrated, interdisciplinary modeling will increase endurance and payload for future UAV systems and allow robust certification of current UAV systems. If this capability is not available in the near future (soon), later introduction of this new capability will be hindered by ad hoc, case-by-case utilization.

The ability to accurately model store carriage and separation will reduce flight test requirements for store certification. This will allow much more rapid analysis of alternative configurations and solutions to specific mission problems.

Improved thermal barrier coatings result in more powerful aircraft with greater readiness and longer service lifetimes. Aircraft with severely spalled coatings are not operational. These improved coatings are dependent on effective, efficient materials' models that promote understanding of bonding mechanisms between the coatings and the bulk material and provide reasonable alternatives to improve this bonding.

Why is HPC needed to address the problem?

HPC resources are currently being extensively used to address issues in all aspects of aerospace vehicle design. For example, hydrocarbon scramjet technology programs use HPC to model engine performance. Significant additional resources are needed to solve the key technical challenges associated with the more ambitious goals of the National Aerospace Initiative (NAI). It is currently impossible to adequately test engine performance at the wide range of flight conditions that are covered by NAI vehicles. Also, structural integrity issues become more critical due to the extended life requirements of reusable vehicles. Computational analysis provides estimates of engine performance and structural integrity:

1. at a fraction of the cost of ground-test and flight-test facilities,
2. at conditions that cannot be achieved in ground-test facilities, and
3. during transients that cannot be achieved in ground-test facilities.

Only HPC systems can adequately model hypersonic propulsion systems to the detail and with the correct physics to accurately predict performance. New HPC capabilities are required to analyze the wide range of possible engine designs and operating parameters necessary to provide good alternatives for future hypersonic vehicles.

Prediction of massively separated flows around aerospace vehicles requires resolution of the unsteady geometry dependent flow features, requiring millions of grid points and thousands of time steps. Currently, only HPC has the resources to tackle such computationally intensive cases. In addition, multiple design alternatives and more operating conditions of the vehicle within its possible flight envelope require large increases in HPC capability to examine possible design trade-offs.

Extreme flight states of UAVs require the highest level of model fidelity and a highly accurate representation of the turbulent flow. To simulate the elastic response of the UAV with nonlinear, unsteady aerodynamics and the most reasonable approximation of turbulence requires HPC of the highest caliber. Providing even more accurate physical models of turbulence and coupling of the fluid flow with the structural models will require capabilities on the orders of magnitude greater than is currently available even with state-of-the art HPC systems.

A major requirement for a detailed CFD or structural mechanics calculation requires an extensive investment in manpower to provide the physical model of the aerospace vehicle over a grid that is used to do the computation. When doing detailed analyses of aircraft-store combinations, a grid of many millions of cells needs to be analyzed for various flight conditions and for each of the various aircraft-store combinations. This results in thousands of analyses that have to be done in a very short amount of time. More automated methods of generating these grids, even possibly re-generating the grid on the HPC system while the calculation is in progress, are required to reduce the total time-to-solution so that timely aircraft-store combinations can be certified.

For accurate materials models, density functional calculations that are generally employed are very computationally intensive. Although lower levels of theory can efficiently treat larger systems, they do not allow the flexibility in the treatment of the physical system. The appropriate physics must be incorporated to ensure that alternative materials proposed for aircraft engine coatings are suitable. Higher levels of theory require extensively greater amounts of HPC capability.

What do you need to accomplish with HPC between 2002 and 2012?

High performance computing can and will make major contributions to the design of aerospace vehicles over the next decade. In engine design, particularly for hypersonic vehicles, we need to provide the design, estimate the performance, and perform the risk-reduction studies of the Single-Engine-Demo scramjet, scheduled for flight in FY 2006. We also need to support the incorporation of three scramjet modules into the NASA X43-C vehicles, scheduled for flight in FY 2007. Additionally, a combined cycle engine demonstrator will fly in FY 2008 or FY 2009. This flight demonstrator will operate from Mach 0.7 through Mach 7. Extensive computational effort will be required in the design and testing of the vehicle and engines as this demonstrator nears flight testing.

As basic turbulence modeling improves to the point where we can accurately and efficiently predict massive flow separations in the air flow around aerospace vehicles, HPC will be used to predict aircraft spin modes, abrupt wing stall under motion, aero-elastic response of aerospace vehicles, unsteady shock/boundary layer interactions at supersonic and hypersonic speeds, and aero-acoustics of cavity flows and jet flows. Greatly increased HPC capability will be required to model these important phenomena in sufficient detail to allow the design of aerospace vehicles based on modeling and simulation.

In addition to the actual computation, high performance computing will also become more important in pre- and post-processing information. Rapid grid generation will allow rapid, efficient set-up of problems in areas such as store separation. Visualization tools will become even more important in interpreting massive amounts of computational data to ensure that maximum utility can be obtained from CFD results.

In materials science, more realistic physical systems must be modeled. These systems will include features such as detailed high temperature dynamics, simulated diffusion and segregation (systems where significant potential barriers must be crossed), and brided length scales to link our detailed atomic-level findings to calculations that can effectively simulate systems that are orders of magnitude larger than those that are feasible with detailed *ab initio* methods. Effective coupling of these length scales from the atomistic to the macroscopic world is particularly challenging and computationally intensive.

What do you need (H/W & S/W) from the HPC Community between 2002 and 2012 to solve your problem?

All of these natural extensions of modeling and simulation for aerospace vehicle design, whether in the dimension of more complete systems or more detailed physics, require orders of magnitude more computing capability. This requirement can best be addressed by large parallel computing machines to handle computationally and memory intensive calculations. In addition, increasingly more capable CFD software, particularly with the ability to handle multi-disciplinary applications (hypersonic flows, grid deformation, chemically reacting flows), is needed. Since much of DoD's large-scale computations are done remotely, and metacomputing is thought to be the wave of the future, increased network transfer rates are also essential to maintaining the ability to efficiently provide large computational resources to a broad community of DoD users. A large amount of on-line data storage is essential, since the results of these calculations, and particularly the large unsteady flow solutions, produce massive amounts of data to be analyzed. For CFD, robust, automated grid generation and adaptation will allow much more efficient use of high performance computing resources and shorten the time-to-solution, particularly for new aerospace vehicle designs. After computational results are produced, flow visualization and analysis tools, particularly for remote users, are essential to make full use of these valuable results.

Project Vision

What is the mission vision?

The overall vision for use of high performance computing for comprehensive design of aerospace vehicles is to develop and use the ability to model, in true interactive fashion, all essential nonlinear aerodynamics, acoustics, electromagnetics, and structural features of current and future aerospace vehicles so that simulations based on these models can have an important impact on their design and operation. Specifically, for engine modeling, DoD researchers are using HPC resources to meet the design, engineering, analytical, and financial challenges of the NAI. This work will unify DoD and NASA researchers, allowing easier interaction and data sharing between team members. In materials modeling, the vision is to provide accurate, detailed materials calculations that can serve as predictive tools for engineering applications and further fundamental scientific understanding of the detailed behavior of heterogeneous interfaces.

What could you accomplish if computers were 10x more powerful?

Depending on the requirements of an analysis, more powerful computers will allow the same solution to be done faster or more accurate solutions to be obtained in the same time. Faster turn-around will allow detailed CFD and structural mechanics studies to flow more easily into the design process. In particular, many more cases of flight conditions for a given design could be investigated. This would be particularly important in planning for a limited number of tests to ensure that the critical points in the test envelope could be investigated experimentally. Also, much more extensive trade-offs in design parameters themselves could be accomplished. More accurate analyses would provide more confidence in the results obtained and allow better understanding of the relevant physical processes.

State-of-the-art turbulence modeling of flows around aerospace vehicles has the feature that as the grid density is increased, more flow features may be resolved as opposed to modeled, reducing modeling errors. Current methods are able to treat full aircraft by modeling the boundary layer, and resolving the flow off the body. If a 10x increase in processor speed was available, more accurate methods could be used to resolve the unsteady flow features in the boundary layer for a portion of the aircraft. This would enable the prediction of the unsteady shock/boundary layer interaction on flaps at supersonic and hypersonic speeds, as well as other problems where unsteady effects in the boundary layer are important to capture. In addition, large increases in computational capability would allow flow simulations to be resolved for all meaningful scales, reducing uncertainty and increasing accuracy. This would make simulation a powerful and cost effective alternative to ground and flight test.

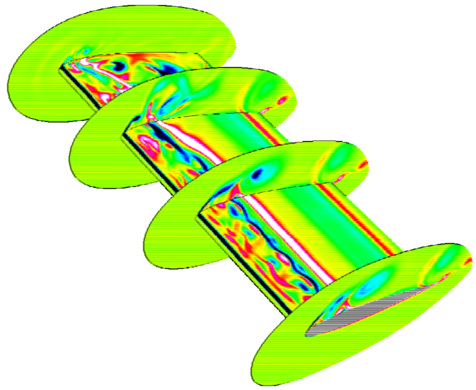
In materials science, large increases in computational power would mean that calculations could be performed that included dynamics for experimentally relevant time scales and temperatures. Likewise, investigations of the effect of low percentage composition dopants could be accomplished more accurately. In addition, we would be able to investigate much larger systems—this could allow simulation of, for example, multi-layered, amorphous, or low symmetry phase materials that are currently problematic. These more relevant calculations would provide much more assurance that proposed new turbine blade coatings based on modeling and simulation would be effective.

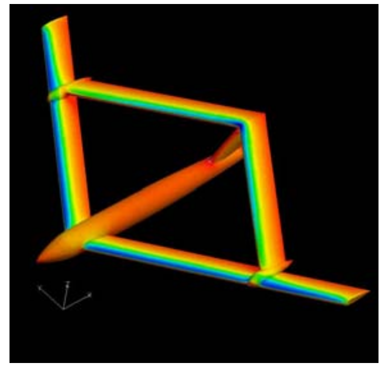
What could you accomplish if computers were 10x cheaper?

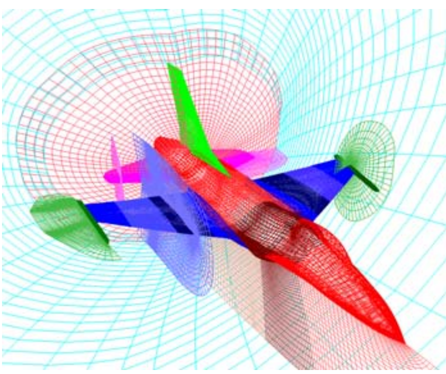
If computational power were significantly cheaper, much larger capabilities and capacity could be acquired for the same budget. Since many of the codes used in this work are scalable, this increased size and/or number of HPC systems could be used to reduce turn-around time dramatically, or increase the problem size treated. This would have essentially the same effect as making HPC systems massively more powerful. High fidelity simulation could move beyond analysis and impact the design cycle, thus avoiding costly redesigns and fixes in engineering design. With many times the computing volume than is currently available, materials scientists could investigate a wider variety of periodic crystals, surfaces, and interfaces. This would allow a more rigorous testing of the validity of predictions before requiring feedback from experimental measures, and thus would result in substantial overall savings as well as faster materials improvements.

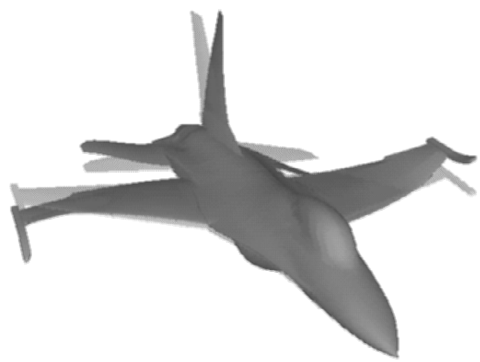
What could you accomplish if computers were 10x easier to program?

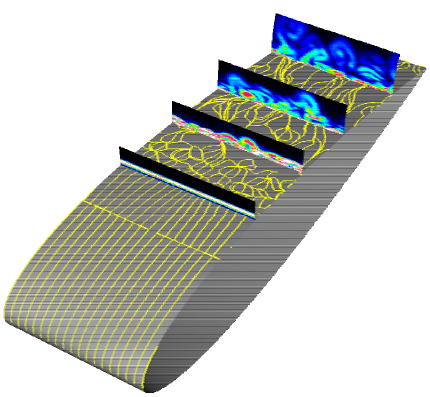
Since a large percentage of analysis time is spent in the process of manipulating input parameters for the software and manipulating output parameters provided by the software, easier programming would mean that more users could develop programs to provide these input/output manipulation routines to meet the needs of their individual analyses, and thus broaden the potential user base for these HPC capabilities. Expanded code features such as deforming grids and real gas effects (for hypersonics) are badly needed for future efforts. An increase in programming efficiency could allow these problems to be handled much sooner than otherwise possible. Currently, each flow solver has a dedicated topology and algorithm. If cost and risk of code development were reduced, algorithmically complex codes could respond dynamically to each application with an optimal strategy. Typical users of HPC systems are engineers and scientists, not computer scientists. Much easier programming models would allow these users to focus far less of their time on programming and much more time on analyzing results, which is what these users are trained to do. New algorithms could be tested in a much more efficient fashion.

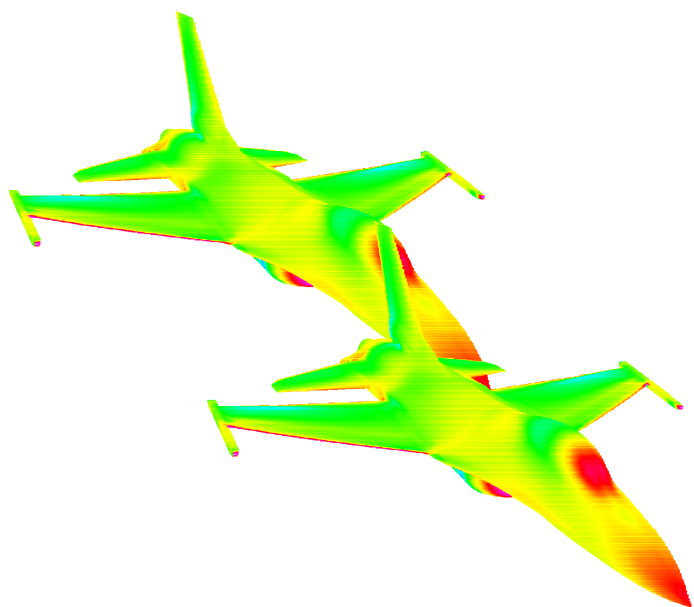












PROJECT NAME: **COMPREHENSIVE AEROSPACE VEHICLE DESIGN**

HPC COMPUTATIONAL REQUIREMENTS

HPC System	Major Software Application	Number of CPU Hours (or sustained Tera op/s) Required for Each Fiscal Year					Security Classification
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012	
IBM SP P3	VASP	500,000	600,000	700,000	750,000	750,000	Unclassified
SP3/Compaq/ Linux	Cobalt	2,000,000	3,000,000	4,000,000	5,000,000	6,000,000	Unclassified
SGI O3K	MM3D	500,000	1,330,000	3,000,000	10,000,000	25,000,000	Top Secret
Linux Clusters	U.S.M3D	1,700,000	2,550,000	3,825,000	5,737,500	8,606,250	Unclassified
IBM SP P3	DNS/LES code	600,000	1,000,000	1,500,000	2,500 K	3,500,000	Unclassified
SGI O3K	Aero-elastic code	200,000	500,000	750,000	1,000,000	1,500,000	Unclassified
Cray SV1	FDL3DI	75,000	100,000	150,000	200,000	200,000	Unclassified
Any	CFD tools	500,000	1,000,000	2,000,000	4,000,000	8,000,000	Secret
Any	CSM tools	100,000	200,000	400,000	800,000	800,000	Secret
Any	CFD tools	500,000	1,000,000	2,000,000	4,000,000	8,000,000	Unclassified
Any	CSM tools	100,000	200,000	400,000	800,000	800,000	Unclassified

NOTES:

A number of CFD tools and CSM tools are to be used. Tool selection depends upon: 1) availability, 2) user experience, and 3) applicability to problem. Since problem scope, and user list is not fully determined at this point, requirements are listed only for the general CFD and CSM areas CFD and CSM tools are also portable to most UNIX-based operating systems. No specific HPC system is required, so requirements are not distributed between various HPC system types

PROJECT NAME: **COMPREHENSIVE AEROSPACE VEHICLE DESIGN**

HPC MEMORY REQUIREMENTS

HPC System	Major Software Application	Typical Memory (GB total per run) Needed Each Fiscal Year				
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012
IBM SP P3	VASP	50	70	70	100	100
SP3/Compaq/Linux	Cobalt	100	200	400	800	1,600
SGI O3K	MM3D	2 GB	10 GB	25 GB	50 GB	100 GB
Linux Clusters	U.S.M3D	40	80	120	200	300
IBM SP P3	DNS/LES code	600 K	1,000 K	1,500 K	2,500 K	3,500 K
SGI O3K	Aero-elastic code	200 K	500 K	750 K	1,000 K	1,500 K
Cray SV1	FDL3DI	75 K	100 K	150 K	200 K	200 K
Any	CFD tools	8	16	32	64	128
Any	CSM tools	1	4	8	16	32
Any	CFD tools	8	16	32	64	128
Any	CSM tools	1	4	8	16	32

Note:

As processor speeds increase, we plan on running larger problems, increasing memory size proportionally. The rates above roughly reflect Moore's law.

PROJECT NAME: **COMPREHENSIVE AEROSPACE VEHICLE DESIGN**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW)

Major Software Application	Computational Technology Area	Short Code Description	Code Size *(Lines of code)*	Rate of Change *(e.g., Fraction of code that changes each year)*	Languages Used *(Fraction of each)*
VASP	CCM	periodic, planewave, pseudo-potential Density Functional Code	~70,000	2%	F90
Cobalt	CFD	Unstructured CFD solver	20,000	30%	f90 (95%), c (5%)
MM3D	CEN	Dense Matrix solution to determine RCS of arbitrary geometry	50K (approximate)	5-10% (estimate)	90% Fortran, 10% C (estimate)
U.S.M3D	CFD	NASA-Langley Finite Volume Unstructured CFD code.	27800	0.05	F90
DNS/LES code	CFD	MPI Parallel, Compact High-Order Code	30,000	5%	F77
Aero-elastic code	CFD+CSM	MPI Parallel Flow Solver with Integrated Linear Structural Solver	20,000	5%	F77 (90%), F90(10%)
FDL3DI	CFD	Vector, Compact High-order Code	25,000	5%	F77

PROJECT NAME: **COMPREHENSIVE AEROSPACE VEHICLE DESIGN**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW)

Major Software Application	Computational Technology Area	Short Code Description	Code Size *(Lines of code)*	Rate of Change *(e.g., Fraction of code that changes each year)*	Languages Used *(Fraction of each)*
CFD tools	CFD	Flow Solver	N/A	10%	Fortran (50%), C (50%)
CSM tools	CSM	Structural Analysis	N/A	10%	Fortran (50%), C (50%)
CFD tools	CFD	Flow Solver	N/A	10%	Fortran (50%), C (50%)
CSM tools	CSM	Structural Analysis	N/A	10%	Fortran (50%), C (50%)
N/A: data not available					
			Key for Computational Technology Area types)		
			Computational Chemistry and Materials Science (CCM) Computational Fluid Dynamics (CFD) Computational Electronics and Nanoelectronics (CEN) Computational Structural Mechanics (CSM)		

PROJECT NAME: **COMPREHENSIVE AEROSPACE VEHICLE DESIGN**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW (CONT.))

Major Software Application	**Programming Models Used**	**Performance Critical Computation** *(e.g., A linear solver or a sorting routine)*	**Time to Solution Requirements** *(S/W development time line requirements)*	**Time to Solution Requirements** *(Problem set-up time line requirements)*	**Time to Solution Requirements** *(Wall clock execution time for specified problem)*	**Time to Solution Requirements** *(Post run analysis time line requirements)*	**Special Characteristics of Code** *(e.g., Is it CPU performance, memory access, or I/O bound)*
VASP	xx	Matrix-Vector, Matrix-Matrix, and 3d FFT	xx	xx	most of the time is execution time	analysis does not require HPC	memory bound
Cobalt	MPI, PVM	fpu operations over an array, future code improvements could be made if a fast sparse matrix solver was available	Most applications are ready to run with the current software. Aero-elastic applications will take about two years to develop the software	two weeks to create a grid	2-4 days for a single solution. 20-40 days for a parametric study (angle of attack sweep)	one day to two weeks depending on the level of detail of the analysis	Unstructured solver, due to the random access pattern for data, is CPU bound, with better performance with faster access to memory, and larger cache size

PROJECT NAME: **COMPREHENSIVE AEROSPACE VEHICLE DESIGN**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW (CONT.))

Major Software Application	Programming Models Used	Performance Critical Computation *(e.g., A linear solver or a sorting routine)*	Time to Solution Requirements *(S/W development time line requirements)*	Time to Solution Requirements *(Problem set-up time line requirements)*	Time to Solution Requirements *(Wall clock execution time for specified problem)*	Time to Solution Requirements *(Post run analysis time line requirements)*	Special Characteristics of Code *(e.g., Is it CPU performance, memory access, or I/O bound)*
U.S.M3D	MPI	CFD Solver	Minimum 10 minutes Maximum 1 hour	1 hour	Minimum 10 minutes Maximum 1 hour	10 minutes	CPU Performance
DNS/LES code	F77 Subroutines	scalar tri- and penta-diagonal linear solvers	3-12 months	2-6 weeks	2-10 days each	1-3 days each	CPU limited
Aero-elastic code	F90 Modules	scalar tri- and penta-diagonal linear solvers	3-12 months	2-6 weeks	1-5 days each	1-3 days each	CPU limited
FDL3DI	F77 Subroutines	scalar tri- and penta-diagonal linear solvers	3-12 months	2-6 weeks	2-10 days each	1-3 days each	SV1 Memory limited
CFD tools	MPI	N/A	0-1 day	1-10 days	1-30 days	1-3 days	N/A
CSM tools	MPI	N/A	0-1 day	1-10 days	1-30 days	1-3 days	N/A
CFD tools	MPI	N/A	0-1 day	1-10 days	1-30 days	1-3 days	N/A
CSM tools	MPI	N/A	0-1 day	1-10 days	1-30 days	1-3 days	N/A
N/A: data not available							

Project Name: **Comprehensive Aerospace Vehicle Design**

HPC Hardware Requirements (Current System Baseline)

HPC System	Number of Computational Nodes Used by Typical Application	Number of Processors Per Node	Processor Characteristics *(e.g., Type, nominal speed)*	Memory Per Processor	I/O System Characteristics	Total Secondary Storage	Interconnect Characteristics *(Local and aggregate bandwidth, latency)*
IBM SP P3	4	16	350 MHz	8GB per 16 CPU's	400 MB/sec	10 GB	—
SP3	16	16	1500Mflop	512Mb	??	??	??
Linux	128	2	933Mflop	512Mb	??	??	??
SGI O3K	2	2	500 MHz, R14000 Chip	1 Gbyte	Fibre Channel	144 Gbytes	
SGI O3K	2	2	500 MHz, R14000 Chip	1 Gbyte	Fibre Channel	144 Gbytes	
SGI O3K	2	2	500 MHz, R14000 Chip	1 Gbyte	Fibre Channel	144 Gbytes	
SGI O3K	2	2	500 MHz, R14000 Chip	1 Gbyte	Fibre Channel	144 Gbytes	
SGI O3K	2	2	500 MHz, R14000 Chip	1 Gbyte	Fibre Channel	144 Gbytes	
SGI O2K	4	2	195 MHz, R10000 Chip	0.5 Gbyte	SCSI 2	108 Gbytes	
SGI O2K	4	2	250 MHz, R10000	0.5 Gbyte	SCSI 2	108 Gbytes	

PROJECT NAME: **COMPREHENSIVE AEROSPACE VEHICLE DESIGN**

HPC HARDWARE REQUIREMENTS (CURRENT SYSTEM BASELINE)

HPC System	Number of Computational Nodes Used by Typical Application	Number of Processors Per Node	Processor Characteristics *(e.g., Type, nominal speed)*	Memory Per Processor	I/O System Characteristics	Total Secondary Storage	Interconnect Characteristics *(Local and aggregate bandwidth, latency)*
			Chip				
SGI O2K	4	2	400 MHz, R12000 Chip	0.5 Gbyte	SCSI 2	108 Gbytes	
SGI O2K	8	2	400 MHz, R12000 Chip	0.25 Gbyte	SCSI 2	108 Gbytes	
Linux Clusters	60	2	Intel PIII/850 MHz	368 Megs	Fast Ethernet	3 Terabytes	Fast Ethernet
IBM SP P3	512	1	xx	xx	xx	xx	xx
SGI O3K	128	1	xx	xx	xx	xx	xx
Cray SV1	4	1	xx	xx	xx	xx	xx
Any	4	4	Compaq Alpha, IBM SP3	1 GB	xx	2 TB	xx

PROJECT NAME: **COMPREHENSIVE AEROSPACE VEHICLE DESIGN**

HPC SOFTWARE REQUIREMENTS (APPLICATION SCALABILITY)

Major Software Application	Scalability of Code for Fixed Problem Size	Scalability of Code for Increasing Problem Size *(Fixed Workload per Processor)*	Efficiency of Code on Target Platform
VASP	~90%	N^3	~90%
Cobalt	Above 95% efficiency as long as 2000 cells per processor. Current runs use about 10M cells, so scalability in theory should be good to 5000 processor for the current problem size. 98% efficiency on 1024 processors for 3M cells has been demonstrated	Over 95% efficiency as long as 2000 cells per processor is maintained.	We generally make sure there are 2000 cells per processor, so we get over 95% efficiency - on all platforms tested (Linux, SP3, T3E)
MM3D	90% CPU used on 8 CPUs, 80-% CPU usage on 32 CPUs	Code maintains constant load per work block per processor; larger problem means more work blocks; Solution time increases dramatically with problem size	98 - 99% CPU utilization on 1 CPU; 80% CPU usage on 32 CPUs
U.S.M3D	> 95%	NA	NA
DNS/LES code	90%	90%	80%
Aero-elastic code	80%	80%	70%
FDL3DI	60%	60%	90%
CFD tools	Fair	Good	Good

PROJECT NAME: **COMPREHENSIVE AEROSPACE VEHICLE DESIGN**			
HPC SOFTWARE REQUIREMENTS (APPLICATION SCALABILITY)			
Major Software Application	**Scalability of Code for Fixed Problem Size**	**Scalability of Code for Increasing Problem Size** *(Fixed Workload per Processor)*	**Efficiency of Code on Target Platform**
CSM tools	N/A	N/A	N/A
CFD tools	Fair	Good	Good
CSM tools	N/A	N/A	N/A
N/A: data not available			

PROJECT NAME: **COMPREHENSIVE AEROSPACE VEHICLE DESIGN**					
HPC SOFTWARE REQUIREMENTS (DETAIL)					
Major Software Application	**Memory Access Pattern** ***(e.g., Stride 1 or random)***	**Computation Profile** ***(e.g., Volume of integer v. floating point computation)***	**I/O Requirements** ***(e.g., Volume and rate)***	**Communication Characteristics** ***(e.g., Describe message size and rate)***	**Inherent Concurrency** ***(e.g., Number of independent threads in main body of calculation)***
VASP	Unknown				
Cobalt	fairly random	Mainly fpu	I/O requirements are generally small compared to CPU requirements. Exception is when unsteady results are being output at a frequent rate.	Domain decomposition with message passing at the zone interfaces. Since the volume of cells scales as N^3, and the interface region as N^2, the ratio of CPU operations to message size scales as N. For large N, communication overhead is small compared to CPU operations.	Single thread for each processor

PROJECT NAME: **COMPREHENSIVE AEROSPACE VEHICLE DESIGN**

HPC SOFTWARE REQUIREMENTS (DETAIL)

Major Software Application	Memory Access Pattern *(e.g., Stride 1 or random)*	Computation Profile *(e.g., Volume of integer v. floating point computation)*	I/O Requirements *(e.g., Volume and rate)*	Communication Characteristics *(e.g., Describe message size and rate)*	Inherent Concurrency *(e.g., Number of independent threads in main body of calculation)*
MM3D		All primary calculations are performed on double precision complex numbers	Currently we have 120 Mbytes/second on a striped dataset over several scsi channels	Typical data package is around 200 Mbytes (varies); data transfers occur between 30 seconds and 4-5 minutes apart	No threads are completely independent; # of threads is equal to # of processors
U.S.M3D	Random	80% Integer/20% FP	1 Gig Total volume More 1 Gig/s IO Rate	NA	1
DNS/LES code	structured, strides	95% FP	2000 MB	50 MB per proc.-time step	1
Aero-elastic code	structured, strides	95% FP	500 MB	30 MB per proc.-time step	2
FDL3DI	structured, strides	95% FP	500 MB		1
CFD tools	N/A	N/A	N/A	N/A	N/A
CSM tools	N/A	N/A	N/A	N/A	N/A
CFD tools	N/A	N/A	N/A	N/A	N/A
CSM tools	N/A	N/A	N/A	N/A	N/A
N/A: data not available					

C-2 SIGNALS INTELLIGENCE

Signals Processing Requirements

+ Classified Application #1

Programming Model used
- + UPC

Performance critical computation (e.g., a linear solver or sorting routing)
- + Signals Processing

Time to solution requirements (wall clock time for a specified problem size)
- + Less than 10 days

Special Characteristics of Code (e.g., CPU memory, or I/O bound)
- + Memory access bound

Computational Technology Area
- + Signals Processing

Code Size (lines of code)
- + 20K lines of code

Rate of Change (e.g., fraction of code that changes each year)
- + 10-20% per year

Languages used (fraction of each)
- + 100% UPC

Scalability of Code for Fixed Size Problem
- + 99.99%

Scalability of Code for Increasing Size Problem (fixed workload per processor)
- + 99.9999%

Efficiency of code on target platform
- + 99.99%

Memory access pattern (e.g., stride-1 vs. random)
- + 70% stride-1, 30% random

Computational profile (e.g., volume of integer vs. floating point computation)
- + 99.99% integer, 0.01% floating point

I/O requirements (e.g., volume and rate)
- + 1GB per hour

Communications characteristics (e.g., describe message size and rate)
- + 1 to 8 word (8 to 64 byte) packets as fast as you can

Inherent concurrency (e.g., number of independent threads in main calculation)
- + several billion threads

Software Applications Code

+ Classified Application #2

Programming Model used
+ Shmem

Performance critical computation (e.g., a linear solver or sorting routing)
+ Signals Processing

Time to solution requirements (wall clock time for a specified problem size)
+ Less than 1 day

Special Characteristics of Code (e.g., CPU memory, or I/O bound)
+ Equal memory access and integer arithmetic operations

Computational Technology Area
+ Signals Processing

Code Size (lines of code)
+ 10K lines of code

Rate of Change (e.g., fraction of code that changes each year)
+ 25% per year

Languages used (fraction of each)
+ 100% C

Scalability of Code for Fixed Size Problem
+ 99.9999%

Scalability of Code for Increasing Size Problem (fixed workload per processor)
+ 99.9999%

Efficiency of code on target platform
+ 99.9999%

Memory access pattern (e.g., stride-1 vs. random)
+ 90% stride-1, 10% random

Computational profile (e.g., volume of integer vs. floating point computation)
+ 99.99% integer, 0.01% floating point

I/O requirements (e.g., volume and rate)
+ 16GB per hour

Communications characteristics (e.g., describe message size and rate)
+ 8KB packets as fast as you can

Inherent concurrency (e.g., number of independent threads in main calculation)
+ ~million threads

Software Applications Code

+ Classified Application #3

Programming Model used
- + UPC, Shmem, and OpenMP

Performance critical computation (e.g., a linear solver or sorting routing)
- + Signals Processing

Time to solution requirements (wall clock time for a specified problem size)
- + Less than 1 day

Special Characteristics of Code (e.g., CPU memory, or I/O bound)
- + equal memory and integer operations

Computational Technology Area
- + Signals Processing

Code Size (lines of code)
- + 5K lines of code

Rate of Change (e.g., fraction of code that changes each year)
- + 50% per year

Languages used (fraction of each)
- + 80% C, 5% FORTRAN, 15% UPC

Scalability of Code for Fixed Size Problem
- + 99.99%

Scalability of Code for Increasing Size Problem (fixed workload per processor)
- + 99.99%

Efficiency of code on target platform
- + 99.99%

Memory access pattern (e.g., stride-1 vs. random)
- + 50% stride-1, 50% random

Computational profile (e.g., volume of integer vs. floating point computation)
- + 99.99% integer, 0.01% floating point

I/O requirements (e.g., volume and rate)
- + almost none

Communications characteristics (e.g., describe message size and rate)
- + 1 word (8 byte) packets or with twice the memory big blocks

Inherent concurrency (e.g., number of independent threads in main calculation)
- + ~million threads

Software Applications Code

+ Classified Application #4

Programming Model used
  + ALL of them

Performance critical computation (e.g., a linear solver or sorting routing)
  + Signals Processing

Time to solution requirements (wall clock time for a specified problem size)
  + Less than 1 day

Special Characteristics of Code (e.g., CPU memory, or I/O bound)
  + one or all depends on problem worked

Computational Technology Area
  + Signals Processing

Code Size (lines of code)
  + 2K lines of code

Rate of Change (e.g., fraction of code that changes each year)
  + changes daily ~300% per year

Languages used (fraction of each)
  + 93% C, 5% FORTRAN, 2% UPC

Scalability of Code for Fixed Size Problem
  + probably very scalable but not written that way due to effort involved and short "lifetime" of code

Scalability of Code for Increasing Size Problem (fixed workload per processor)
  + probably very scalable but not written that way due to effort involved and short "lifetime" of code

Efficiency of code on target platform
  + probably very scalable but not written that way due to effort involved and short "lifetime" of code

Memory access pattern (e.g., stride-1 vs. random)
  + 70% stride-1, 30% random but sometimes 50% stride-1, 50% random

Computational profile (e.g., volume of integer vs. floating point computation)
  + 80% integer, 20% floating point

I/O requirements (e.g., volume and rate)
  + very substantial

Communications characteristics (e.g., describe message size and rate)
  + 1 word (8 byte) packets as fast as you can

Inherent concurrency (e.g., number of independent threads in main calculation)
  + several thousand to millions but we don't write it that way

Project Information

HPC Project Name: **Signals Analysis**

Agency and Organization: National Security Agency, SIGINT Programs Office, FT. George G. Meade, MD 20755

Number of Users in Project: Hundreds

Computational Technology Area(s) (See attached list): Graphical Computations

Project Description

A massive Knowledge Management system that will advance information technologies to enable aggregate and relate thousands of information islands that exist.

What is the real life problem?

Need a means to aggregate disparate types of information.

Need information together in a single, computationally complete, logically consistent form on a massive scale.

What is the National Security Impact of (solving and not solving) the problem?

Better prediction in intelligence analysis.

Better response (less than a minute) than the current several days to weeks on information queries.

Increased production for future intelligence product reporting.

Why is HPC needed to address the problem?

Multiple computationally intensive calculations involving graphs.

Massive amounts of information and large amounts of memory are required for operational results.

Billions of computational graphical nodes are required.

What do you need to accomplish with HPC between 2002 and 2012?

Predictions in intelligence analysis

Information layer interfaces

Complete algorithm development

Apply complex graph configurations

What do you need (H/W & S/W) from the HPC Community between 2002 and 2012 to solve your problem?

Large parallel compute machines

Increased memory and capacity in super computers

Increased transfer rates in all stages of network (to / from subsystem I/O centers)

Project Vision

What is the mission vision?

Instantaneous, real-time response to a 20 plus petabyte of data.

What could you accomplish if computers were 10x more powerful?

A 50% increase in predictive analysis product reporting with more processing and memory.

What could you accomplish if computers were 10x cheaper?

A 25% increase in predictive analysis product reporting.

What could you accomplish if computers were 10x easier to program?

(All three of the above questions to be answered relative to progress greater than that expected from Moore's Law)

A 10% increase in predictive analysis and product reporting.

Project Name: **Signals Intelligence**

HPC Computational Requirements

HPC System	Major Software Application	Number of CPU Hours (or sustained Tera op/s) Required for Each Fiscal Year					Security Classification
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012	
SUN E-15K	GPC	128G	170G	212G	245G	310G	FOUO

Project Name: **Signals Intelligence**

HPC Memory Requirements

HPC System	Major Software Application	Typical Memory (GB total per run) Needed Each Fiscal Year				
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012
SUN E-15K	GPC	240	432	624	816	1842

PROJECT NAME: **SIGNALS INTELLIGENCE**					
HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW)					
Major Software Application	**Computational Technology Area**	**Short Code Description**	**Code Size** *(Lines of code)*	**Rate of Change** *(e.g., Fraction of code that changes each year)*	**Languages Used** *(Fraction of each)*
GPC	Graph Theory	N/A	<30K	<25%	
				Key (for Computational Technology Area types)	
				Graph Theory; Combinatorics	

PROJECT NAME: **SIGNALS INTELLIGENCE**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW (CONT.))

Major Software Application	Programming Models Used	Performance Critical Computation *(e.g., A linear solver or a sorting routine)*	Time to Solution Requirements *(S/W development time line requirements)*	Time to Solution Requirements *(Problem set-up time line requirements)*	Time to Solution Requirements *(Wall clock execution time for specified problem)*	Time to Solution Requirements *(Post run analysis time line requirements)*	Special Characteristics of Code *(e.g., Is it CPU performance, memory access, or I/O bound)*
GPC		linear	<1hr.	<1hr.	<1min.	<1min.	memory

PROJECT NAME: **SIGNALS INTELLIGENCE**							
HPC HARDWARE REQUIREMENTS (CURRENT SYSTEM BASELINE)							
HPC System	**Number of Computational Nodes Used by Typical Application**	**Number of Processors Per Node**	**Processor Characteristics** *(e.g., Type, nominal speed)*	**Memory Per Processor**	**I/O System Characteristics**	**Total Secondary Storage**	**Interconnect Characteristics** *(Local and aggregate bandwidth, latency)*
SUN E-15K	16	24	U.S. III/900 MHZ	8 GB	1 GB per Second	160 GB	GIGE

PROJECT NAME: **SIGNALS INTELLIGENCE**

HPC SOFTWARE REQUIREMENTS (APPLICATION SCALABILITY)

Major Software Application	Scalability of Code for Fixed Problem Size	Scalability of Code for Increasing Problem Size *(Fixed Workload per Processor)*	Efficiency of code on target platform
GPC	Unknown at this time	Unknown at this time	Unknown at this time

PROJECT NAME: **SIGNALS INTELLIGENCE**					
HPC SOFTWARE REQUIREMENTS (DETAIL)					
Major Software Application	**Memory Access Pattern** *(e.g., Stride 1 or random)*	**Computation Profile** *(e.g., Volume of integer v. floating point computation)*	**I/O Requirements** *(e.g., Volume and rate)*	**Communication Characteristics** *(e.g., Describe message size and rate)*	**Inherent Concurrency** *(e.g., Number of independent threads in main body of calculation)*
GPC	random	integer - 43G	1 GB per Second	< 1 GB per second	16

C-3 OPERATIONAL WEATHER/OCEAN FORECASTING

HPC Project Name: **Operational Weather/Ocean Forecasting**

Agency and Organization: Naval Meteorology and Oceanography Command, Fleet Numerical Meteorology and Oceanography Center, Naval Research Laboratory, Monterey, CA 93943

Project Leader(s): a) Dr. William Burnett, b) Mr. Mike Clancy, c) Dr. Richard Hodur, and d) Dr. Alan Wallcraft

Project Leader E-Mail Address: a) burnettb@cnmoc.navy.mil, b) mike.clancy@fnmoc.navy.mil, c) hodur@nrlmry.navy.mil, and d) wallcraft@nrlssc.navy.mil

Project Leader Phone Number: a) 228-688-4766, b) 831-656-4414, c) 831-656-4788, and d) 228-688-4813

Project Leader Address: a) 1100 Balch Blvd, Stennis Space Center, MS 39529-5001, b) 7 Grace Hopper Ave, Stop 1, Monterey, CA 93943-5501, c) 7 Grace Hopper Ave, Monterey, CA 93943-5502, and d) NRL Code 7323, Stennis Space Center, MS 39529

Number of Users in Project: 250 Project Collaborators

Computational Technology Area(s) (See list in companion spreadsheet, sheet 3): CWO

Project Description

The Primary Oceanographic Prediction System (POPS) produces and provides critical, classified and unclassified atmospheric and oceanographic guidance to Navy and Department of Defense (DoD) activities worldwide on a fixed schedule, 24 hours a day. POPS covers the entire Fleet Numerical Meteorology and Oceanography Center (FNMOC) enterprise including the supercomputing, communications (including receipt of hundreds of thousands of observations and transmission of model products), databases, data assimilation and distribution, and systems control/monitoring. The operations require a large investment in equipment that depreciates rapidly. POPS is the engine that operates all the Navy's global/regional/tactical atmospheric, oceanographic, wave, ice, and tropical cyclone models, and DoD's only coupled air/ocean model.

What is the real life problem?

The POPS is the only national system that assimilates classified and unclassified data, and produces and disseminates classified and unclassified global/regional atmospheric guidance that is used by:

the Navy to operate their global ocean, regional and tactical ocean/atmosphere/wave/ice/tropical cyclone models from the unclassified to the SCI level, and their eight distributed tactical forecast systems,

the Air Force to operate their regional atmospheric models, cloud prediction systems, and strategic decision aids as specified via the Navy/Air Force agreement

the Joint Forces Command, Defense Threat Reduction Agency and Lawrence Livermore National Laboratory to operate their Weapons of Mass Destruction decision aids and to aid in contingency planning, the Central Intelligence Agency to aid in their contingency planning,

the national backup to the National Weather Service's supercomputer, there is no global backup for the nation,

the U.S.STRATCOM for ballistic missile support,

the Presidential Support Unit to aid in their contingency planning.

Generating numerical weather/ocean forecasts requires the assimilation of in-situ and remotely sensed observations into a 3-D analysis of the atmosphere that is used as initial conditions for a numerical weather prediction model. The assimilation of the observations requires the solution of complex linear algebra and variational equations, and the model requires the numerical solution of a system of non-linear, 3-D, partial differential equations, and the computation of complex physical processes, all of which requires extensive computational resources.

What is the National Security Impact of (solving and not solving) the problem?

A lack of enhanced capabilities will have a ripple effect within a number of areas in the Navy and DoD and will precipitate a crisis due to the lack of weather/ocean numerical guidance. A similar incident occurred in 1999 when the National Weather Service's supercomputer system caught on fire and failed during the crucial hurricane season when a number of storms were off the U.S. coast. The POPS system served as a contingency backup to the National Weather Service until a replacement system was brought on line three months later.

Inadequate high-end operational supercomputing resources in DoD have hampered the ability to continue providing real-time weather and coupled modeling forecasts to the warfighter. The POPS will soon be outdated, with the attendant degradation and ultimate loss of the capability to predict weather in various operating environments. This is especially critical during a time when Enduring Freedom, Homeland Security, and the War on Terrorism are placing a demand for very-high resolution atmospheric and oceanographic models that are able to calculate the dispersion of air/water-borne threats to the Fleet and the nation. POPS will not be able to provide tailored high-resolution oceanographic and atmospheric prediction and dispersion products that will address the full range of requirements and threats facing the nation daily. The POPS will not be able to operate the R&D models that are scheduled for transition in the FY 04–FY 09 timeframe, including: aerosol/chemical dispersion forecasts, target area weather predictions, support to on-scene modeling, improved tropical cyclone forecasts, and very-high atmosphere/space forecasts.

Why is HPC needed to address the problem?

Numerical weather/ocean prediction requires the solution of a complex system of equations (e.g., Newton's Second Law of Motion, 1st Law of Thermodynamics) that can only be solved numerically, using large computer systems with fast processing capabilities and fast memory access.

Large increases in observational data (particularly satellites) require the inversion of very large matrices in data assimilation systems.

Increases in model resolution and inclusion of important physics are directly tied to increases in computational power and speed.

“Doubling” the model resolution requires 16 times as many computations (x, y, z, and time).

Current state-of-the-art support to operational users is computationally constrained; improvements to models and systems supporting the on-scene commander are even more so.

Ensemble modeling helps quantify the certainty of a forecast, however this comes at the price of making many additional numerical forecasts (typically 10–50 more forecasts).

Numerical weather/ocean prediction is a repeatable process; it must be performed whenever new data is available, typically every 6-12 hours every day of the year.

Ocean prediction requires 20 to 30 times smaller than atmospheric phenomena – and requires about 3-4 km grid resolution.

First generation ocean prediction products very aggressively minimized computational cost, at the expense of some capability:

Initial system was 7 vertical layers with 7 km grid resolution—216 IBM WinterHawk II processors

Follow on systems will need to run a 3-4 km model with more resolution in the vertical. Future systems will operate a coupled ocean, wave, ice global model.

What do you need to accomplish with HPC between 2002 and 2012?

Meet the warfighter modeling requirements stated in the Administrative Model Oversight Panel (AMOP) modeling Roadmap: https://www.cnmoc.navy.mil/nmosw/staff/roadmap/cnmocweb/webpages/cnmochome.html.

Ability to process/assimilate expected large increases of remotely sensed data using advanced data assimilation methods (e.g., 4-D variational assimilation and Kalman filtering).

Demonstrate the ability to make reliable numerical predictions, including aerosol transport, on the micro-alpha scale (0.2-2 km) for urban-sized areas.

Increased resolution for global and mesoscale models, improved data assimilation systems, quality control, analysis techniques, coupling of ocean/atmosphere models, and inclusion of aerosols within urban environments.

Increase the number of geographical areas to cover with high-resolution models.

Improve timeliness to provide the warfighter with real-time weather/ocean products.

What do you need (H/W & S/W) from the HPC Community between 2002 and 2012 to solve your problem?

Hardware (size and speed)

Increased CPU speed and memory access times

Sustained computational processing rate and improved compiler technology

More commonality between HPC sites (batch queue systems, file structures, user names, etc).

Improved visualization techniques

Network speed and bandwidth for transferring files between centers

Increased disk storage – increased model resolution requires 8x increase in storage requirements

Currently limited to about 500 processors per job. Partly because the largest DoD machines have about 1,300 CPUs but also because of interconnect latency. Would like to see 5,000 processor machines with much lower latency interconnect, allowing the use of 1,000 to 2,000 processors per job. The interconnect should be optimized for one-sided (direct to memory) communications

Software (operations specific optimization)

Multi-level security

Operations-run optimization software improvements – keeps wall clock time commensurate with sustained and peak speeds available.

Faster data assimilation systems – large increases in remotely sensed data redundancy present new problem types

Automatic adjoint code development software

Load balancing

Job sharing and triggering

Shared file system improvements

MPI (i.e., message passing) is limiting performance. Would like to see portable and standardized software for one-sided communications, e.g., SHMEM or Co-Array Fortran

Project Vision

What is the mission vision?

Detailed knowledge of the atmosphere and the ocean to ensure the warfighters minimize risk and maximize combat effectiveness.

High-resolution (9 km or less) air, ocean, land, ice, wave information from a global prediction system for 0-14 days.

Analysis and prediction (0-3 days) of the environment in urban areas with horizontal resolutions of less than 1 km.

Aerosol and chemical-biological support (transport and dispersion) included in all numerical weather predictions.

What could you accomplish if computers were 10x more powerful?

This implies that the interconnect latency between processors has been drastically reduced, allowing us to efficiently use either 5,000 processors, or we use 500 processors that are 10x faster. Achieving 10x faster processors would be equivalent to approximately 5 years of Moore's Law increase in performance.

It would bring forward the earliest date any given operational system would be deployed, in principle (i.e., the earliest date that computers are fast enough to run the product) by 5 years. We schedule delivery of new operational products very close to this "earliest practical" date, but a sudden increase of 10x might only bring forward delivery of the next product by 3 years because the pre-existing (10x slower) computers would not be capable of doing the necessary R&D in time.

Having stated that fact, there is a possibility that the weather/ocean modeling roadmap could be advanced roughly 3–5 years. This would allow operational implementation of the 4-D variational data assimilation systems that are very intensive computationally, the global atmospheric model could be operated at 25 km resolution, the global ocean model could be operated at 7 km resolution, and very-high resolution numerical weather/ocean models could begin forecasting for short-term chemical/biological dispersion events with high accuracy.

What could you accomplish if computers were 10x cheaper?

If they were cheaper, but not more capable (i.e., had existing interconnect latency), then total capacity (throughput) would be increased 10x but we could not use more than about 500 processors per job because of MPI communication overhead. This would greatly speed up the development of future operational ocean prediction systems, and allow them to be much better tested, but it would not allow a higher resolution global ocean prediction system to run operationally. This is because of the near real-time constraints on operational forecasts. It might allow ensembles (multiple related predictions) to be used operationally. For example, if one forecast took one wall hour on 500 processors then we could run 10 related forecasts in two hours on 2500 processors (or five 500-processor machines).

Note that cheaper with higher latency interconnect already exists (Beowulf cluster) and is not useful for ocean prediction because such systems don't scale to O(100) processors on ocean models.

What could you accomplish if computers were 10x easier to program?

(All three of the above questions to be answered relative to progress greater than that expected from Moore's Law)

A portable one-sided API would simplify programming, but perhaps only by 3x. It would allow us to completely remove the complicated and machine specific code required to make MPI efficient on a range of machines.

If computers were 10x easier to program, with no loss of efficiency, then it would be easier to explore alternative ocean model designs. This might lead to a breakthrough in ocean model performance, but this cannot be guaranteed. In general, it is more likely to get a 10x improvement in performance from software (new algorithms) than hardware (Moore's Law).

PROJECT NAME: **OPERATIONAL WEATHER/OCEAN FORECASTING**

HPC COMPUTATIONAL REQUIREMENTS

HPC System	Major Software Application	Number of CPU Hours (or sustained Tera flops) Required for Each Fiscal Year					Security Classification
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012	
POPS	Global Weather Modeling	0.1	2	4	12	24	Multi-Level Security (Unclassified, Class, SCI)
	Mesoscale Modeling	0.1	2	4	12	24	Multi-Level Security
	Pre& Post Processing and Visualization	*160 1.5 GHz CPUs	* 3200 1.5 GHz CPUs	*6400 1.5 GHz CPUs	*20000 1.5 GHz CPUs	*40000 1.5 GHz CPUs	Require U, C, S, TS, and Certified Communication Bridges
IBM SP3	Global Ocean Models	0.5	2	8	64	256	Unclassified

POPS
Primary Oceanographic Prediction System - SGI O3K 512 Processors in FY02.
* POPS (Pre&Post Processing and Visualization) performance requirements are stated as numbers of required 1.5 GHz CPUs. Of course, Moore's Law will reduce the actual numbers of CPUs required in the out years.

PROJECT NAME: **OPERATIONAL WEATHER/OCEAN FORECASTING**

HPC MEMORY REQUIREMENTS

HPC System	Major Software Application	Typical Memory (GB total per run) Needed Each Fiscal Year				
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012
POPS	Global Weather Modeling	0.5	4	8	24	48
	Mesoscale Modeling	0.5	4	8	24	48
	Pre& Post Processing and Visualization	0.5	4.3	8.6	26	52
IBM SP3	Global Ocean Models	100	200	400	1200	2400

POPS
Primary Oceanographic Prediction System - SGI O3K 512 Processors in FY02
 * POPS (Pre&Post Processing and Visualization) performance requirements are stated as numbers of required 1.5 GHz CPUs. Of course, Moore's Law will reduce the actual numbers of CPUs required in the out years.

PROJECT NAME: **OPERATIONAL WEATHER/OCEAN FORECASTING**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW)

Major Software Application	Computational Technology Area	Short Code Description	Code Size *(Lines of code)*	Rate of Change *(e.g., Fraction of code that changes each year)*	Languages Used *(Fraction of each)*
Global Weather Modeling	CWO	Spectral Numerical Weather Prediction	430000, *630000 (after next major upgrade in 2003)	Varies Greately, usually < 10%, Occnly > 50% for major upgrades	Fortran
Mesoscale Modeling	CWO	Nested-Grid Numerical Weather Prediction	706,000	Varies Greately, usually < 10%, Occnly > 50% for major upgrades	Fortran
Pre& Post Processing and Visualization	CWO	Data Ingest, Product Generation, Data/Product Distribution, Including Web Services	3,575,000	Varies Greatly Usually <10%, Occnly >50% as new data types emerge and new products are generated	C/C++ 40% ADA 10% FORTRAN 5% Remainder is other: PERL, JAVA, HTML, XML, KORN SHELL, IDL, ETC…
Global Ocean Models	CWO	Primitive Equation Ocean Model	250,000	10%-50%	Fortran

Primary Oceanographic Prediction System - SGI O3K 512 Processors in FY02
* POPS (Pre&Post Processing and Visualization) performance requirements are stated as numbers of required 1.5 GHz CPUs.
Of course, Moore's Law will reduce the actual numbers of CPUs required in the out years.

			Key (for Computational Technology Area types)
			Climate/Weather/Ocean Modeling and Simulation (CWO)

PROJECT NAME: **OPERATIONAL WEATHER/OCEAN FORECASTING**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW (CONT.))

Major Software Application	Programming models used	Performance critical computation (e.g., A linear solver or a sorting routine)	Time to solution requirements (S/W Development time line requirements)	Time to solution requirements (Problem set-up time line requirements)	Time to solution requirements (Wall clock execution time for specified problem)	Time to solution requirements (Post run analysis time line requirements)	Special Characteristics of Code (e.g., Is it CPU performance, memory access, or I/O bound)
Global Weather Modeling	MPI	Matrix Inversion			8 min per forecast day	xx	CPU Perf Memory Access
Mesoscale Modeling	MPI	Solution of System of Non-Linear PDE's		xx	12 min per forecast day	xx	CPU Perf Memory Access
Pre& Post Processing and Visualization	Beowulf Clusters, Batch, Event Driven, CRON, Multi-Tiered E-Business	Data Ingest and Data Manipulation	xx	xx	Product Request to Product Receipt: 80% < 2 mins; 100% < 5 Mins	xx	CPU Perf Memory Access IO
Global Ocean Models	MPI, OpenMP, SHMEM	Explicit Finite Difference			10 min per forecast day	xx	Memory Access and Communications Latency

Primary Oceanographic Prediction System - SGI O3K 512 Processors in FY02

* POPS (Pre&Post Processing and Visualization) performance requirements are stated as numbers of required 1.5 GHz CPUs.

Of course, Moore's Law will reduce the actual numbers of CPUs required in the out years.

PROJECT NAME: **OPERATIONAL WEATHER/OCEAN FORECASTING**

HPC HARDWARE REQUIREMENTS (CURRENT SYSTEM BASELINE)

HPC System	Number of Computational Nodes Used by Typical Application	Number of Processors Per Node	Processor characteristics *(e.g., Type, nominal speed)*	Memory Per Processor	I/O System Characteristics	Total Secondary Storage	Interconnect Characteristics *(Local and aggregate bandwidth, latency)*
Global/Mesoscale Weather Models	30 for Global Modeling 20 Per Model Domain for Mesoscale Modeling	4	SGI O3K 1 Gflop	1 GB	High-speed fiber channel San, 8 TB Raid, 100 MB/S BW per connection, *HIPPI Connection Between AMS and ATOS, 800 MB/S	80 TB Nearline Storage in Automated Type Silo System	100 MB/S BW per connection for Raid Disk
Pre and Post Processing and Visualization	Varies, Some Clustered, Some Single CPU Batch	Varies, 18 for E10K, 1 for 30-Node Cluster	Varies, 333 MHz Sun to 1.2 GHz Intel	Varies 0.5 GB/CPU to 1GB/CPU	Varies, *HIPPI (as Above) GIGE 2.7 TB SAN	Total for all subsystems: 3TB of Near-line Tape Storage	Varies
Global Ocean Model	125	4	IBM WinterHawk II (1.5 Gflops)	1 GB	IBM GPFS filesystem	100 TB Archive	HIPPI Interconnect

PROJECT NAME: **OPERATIONAL WEATHER/OCEAN FORECASTING**

HPC SOFTWARE REQUIREMENTS (APPLICATION SCALABILITY)

Major Software Application	Scalability of Code for Fixed Problem Size	Scalability of Code for Increasing Problem Size *(Fixed Workload per Processor)*	Efficiency of Code on Target Platform
Global Weather Modeling	Linear to ~200 Processors	Linear as Long as Memory Increases Propotionally	N/A Target Platform not Identified
Mesoscale Modeling	Linear to ~200 Processors	Linear as Long as Memory Increases Propotionally	N/A Target Platform not Identified
Pre& Post Processing and Visualization	Varies, Some applications with limited scalability, but many applications embarrassingly parallel	N/A	N/A Target Platform not Identified
Global Ocean Models	Linear to ~200 Processors		~10% (sustained/peak flops)

Primary Oceanographic Prediction System - SGI O3K 512 Processors in FY02
* POPS (Pre&Post Processing and Visualization) performance requirements are stated as numbers of required 1.5 GHz CPUs. Of course, Moore's Law will reduce the actual numbers of CPUs required in the out years.

Project Name: **OPERATIONAL WEATHER/OCEAN FORECASTING**

HPC SOFTWARE REQUIREMENTS (DETAIL)

Major Software Application	Memory Access Pattern *(e.g., Stride 1 or random)*	Computation Profile *(e.g., Volume of integer v. floating point computation)*	I/O Requirements *(e.g., Volume and rate)*	Communication Characteristics *(e.g., Describe message size and rate)*	Inherent Concurrency *(e.g., Number of independent threads in main body of calculation)*
Global Weather Modeling	Both	FP Computional Dominated	1 TB per 12 Hours (Combined global and Mesoscale)	Varies Widely	xx
Mesoscale Modeling	Both	FP Computional Dominated	1 TB per 12 Hours (Combined global and Mesoscale)	Varies Widely	xx
Pre& Post Processing and Visualization			Average 2003 72 GB/Day 2004 800 GB/Day 2006 5000 GB/Day Peak 2003 14 GB/Hr 2004 28 GB/HR 2007 42 GB/HR	Varies Widely, Small # of HI Volume Customers, Large # of Low Volume Customers, Ingest 10^7 Observations per day	xx
Global Ocean Models	Stride-1	FP Computional Dominated	0.5 TB per day	Low latency required 2-d domain decomposition over 9000x6000 domain, so O(20,000) with low enough latency	xx

Primary Oceanographic Prediction System - SGI O3K 512 Processors in FY02
* POPS (Pre&Post Processing and Visualization) performance requirements are stated as numbers of required 1.5 GHz CPUs.
Of course, Moore's Law will reduce the actual numbers of CPUs required in the out years.

C-4 STEALTHY SHIP DESIGN

Project Information

HPC Project Name: Stealthy Ship Design

Agency and Organization: Office of Naval Research, 800 N. Quincy Street, Arlington, VA 22217

Project Leader(s): a) Dr. Ki-Han Kim, b) Dr. Joseph J. Gorski, c) Dr. Douglas Dommermuth, d) Dr. Lafe Taylor, e) Dr. Robert V. Wilson, f) Prof. Dick K.P. Yue, and g) Dr. Mark Hyman

Project Leader E-Mail Address: a) kimk@onr.navy.mil, b) GorskiJJ@nswccd.navy.mil, c) dommer@trg.saic.com, d) lafe@erc.msstate.edu, e) robert-wilson@uiowa.edu, f) yue@mit.edu, and g) HymanMC@ncsc.navy.mil

Project Leader Address: a) Office of Naval Research, Code 333, 800 N. Quincy Street, Arlington, VA, b) Naval Surface Warfare Center, Carderock Division, Code 5400, 9500 MacArthur Boulevard, West Bethesda, MD 20817-5700, c) Science Applications International Corp., Naval Hydrodynamics Division, 10260 Campus Point Drive, M/S C4, San Diego, CA 92121, d) NSF Engineering Research Center, P.O.Box 9627, Mississippi State University, Starkville, MS 39762, e) Iowa Institute of Hydraulic Research, The University of Iowa, Iowa City, Iowa 52242-1585, f) Department of Ocean Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, and g) Naval Surface Warfare Center, Coastal Systems Station, Panama City, FL 32407-7001

Project Leader Phone: a) 703-696-4305, b) 301-227-1930, c) 858-826-6911, d) 662-325-7299, e) 319-384-0684, f) 617-253-6823, and g) 850-234-4126

Number of Users in Project: 30

Computational Technology Area(s) (See list in companion spreadsheet, sheet 3): Computational Fluid Dynamics

Project Description

The objective of this project is to develop and demonstrate a simulation-based design environment for future U.S. Navy advanced stealthy surface ships using Computational Fluid Dynamics (CFD) tools with minimal empiricism. Four state-of-the-art CFD codes, two unsteady Reynolds-averaged Navier Stokes codes (UNCLE and CFDSHIP-IOWA), and two large-eddy simulation (LES) codes (ShipLES and NFA), are used to predict the fully nonlinear turbulent free-surface flow around surface ships. These physics-based hydrodynamic simulation tools will be combined with a design optimization process to efficiently explore new ship hull forms that are drastically different from traditional hull forms.

What is the real life problem?

Naval ships of the future must meet emerging missions and related operational requirements, and accommodate emerging technologies such as electric drive as the main propulsion system. Emerging missions include increased littoral operations that would require unprecedented ship signature reduction. To meet these requirements, future ships will be radically different from those currently in the fleet. Current design/analysis methodology, primarily based on potential-flow theory and empiricism, cannot design or predict hydrodynamic ship signatures adequately due to lack of proper flow-physics modeling. Additionally, the historical database does not allow assessment of the seakeeping and maneuvering behavior of radically new hull forms. The U.S. Navy needs a way to analyze these new hull forms to better understand the flow physics and associated operating behavior in order to design the next generation of surface combatants.

What is the National Security Impact of (solving and not solving) the problem?

The ORD (Operational Requirement Document) of the DD-21 (now DD(X) program) land-attack surface combatant requires ship signature reduction to levels comparable to submarine signatures. Providing the new computational capability will significantly reduce the developmental cost and design-cycle time for the U.S. Navy's future stealthy ships, and improve the chances of success. Without this capability, signature mitigation requires a traditional build-and-test approach that could prove to be prohibitively costly and time-consuming.

Why is HPC needed to address the problem?

Prediction of high-Reynolds number turbulent flows to simulate large-scale waves and ship motions is only possible through HPC systems. A very wide range of scales must be resolved in order to ensure accuracy. Large computations demanding massively parallel machines are essential. The simulation of turbulent wakes, steep breaking waves, the entrainment of air, and the generation of spray using unsteady RANS and LES codes are computationally intensive because of the turbulent nature of the flow and because of the fidelity necessary to resolve complex free-surface interactions. Such simulations are feasible only on HPC systems.

What do you need to accomplish with HPC between 2002 and 2012?

Current prediction capability needs to be extended to coupled 6-DOF motions in seaways and to accurately simulate the turbulent wakes, steep breaking waves, air entrainment, and the generation of spray for a full scale surface combatant. A series of validations from model- to large- to full-scale needs be made to assess the accuracy of the modeling of the complex flow physics, thus improving the prediction capability. Such calculations are computationally intensive because of the turbulent nature of the flow and because of the fidelity necessary to resolve complex free-surface interactions. Additionally, such computations need to be done much faster than feasible today in order to do design trade-offs and hull form optimization in an acceptable time frame to impact design decisions.

What do you need (H/W & S/W) from the HPC Community between 2002 and 2012 to solve your problem?

To achieve the above goals, the hardware should improve in speed on the order of 20-30 times current capabilities. Additionally, the number of processors available for the computations needs to be significantly increased. This is because the computations for a particular hull form are extremely demanding. To do a single maneuver takes on the order of 20,000+ CPU hours on a current Cray T3E. To get this down to an acceptable turn around time for design iterations will require both faster processors and more of them to spread out the computations.

Although the U.S. Navy has been investing in the development of software primarily at the universities and some selected industrial partners, there is no established mechanism that warrants a long-term development, support, and maintenance of computer codes. Consequently, commercial CFD codes are of increasing interest to the U.S. Navy, such as CFD++ and FLUENT. We will need some promising commercial software maintained on HPCMO computers. Separately, to deal with these large data sets we will need an efficient means of post-processing the data, and visualizing the necessary parts of it on the HPCMO machines, as the data sets will be far too large to regularly bring back to the PIs' local sites.

Project Vision

What is the mission vision?

Over the next ten years this project will develop a complete suite of computational tools to quantify the myriad physics associated with mixed-phase turbulent flow in the near field flow of a ship. This computational capability will be developed to the point where it can be used for trade-off and optimization studies in a timely enough manner to significantly impact and drive the development of future stealthy ships. Such a computational capability can help radically transform our future fleet so that it can deal with emerging new threats rather than incrementally changing our current fleet with current technology.

What could you accomplish if computers were 10x more powerful?

With 10x more powerful computers we could begin to accurately predict a number of complex flow physics events important to stealthy ship design including: unsteady motions and waves, near- and far-field turbulent and surface-wave wakes, transom-stern separation, spilling breaking waves, air entrainment, separation along the contact line, capturing of spray sheets and parameterization of spray sheets, and extreme events such as green water on deck.

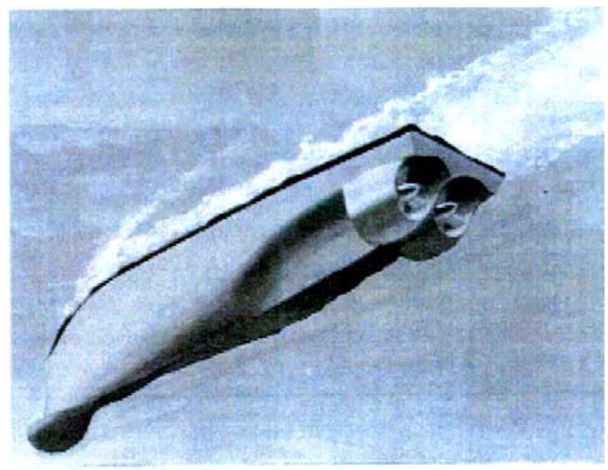
What could you accomplish if computers were 10x cheaper?

Assuming this would lead to larger numbers of processors available this would lead to being able to do the above computations in a timely manner that design trade-offs and optimization studies could be performed.

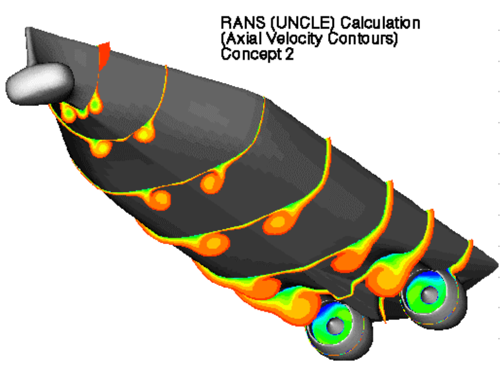
What could you accomplish if computers were 10x easier to program?

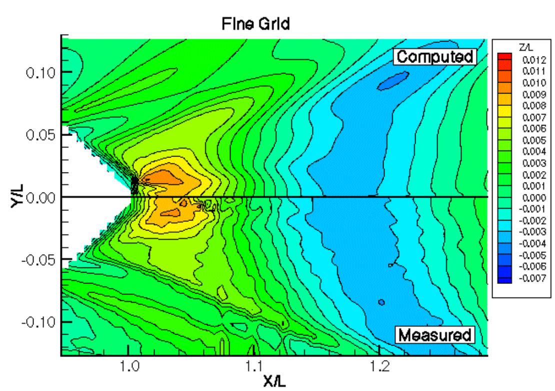
(All three of the above questions to be answered relative to progress greater than that expected from Moore's Law)

It would allow the above to be accomplished a little faster, but is not as important as the other two particularly as we move to using commercial codes.

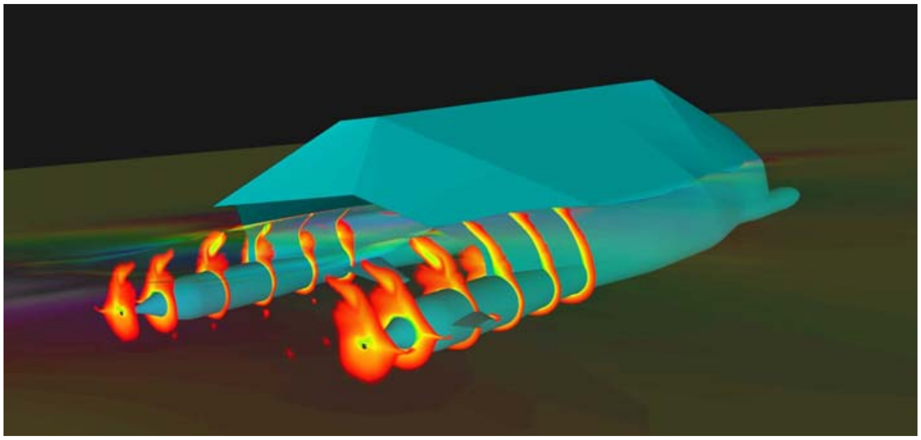


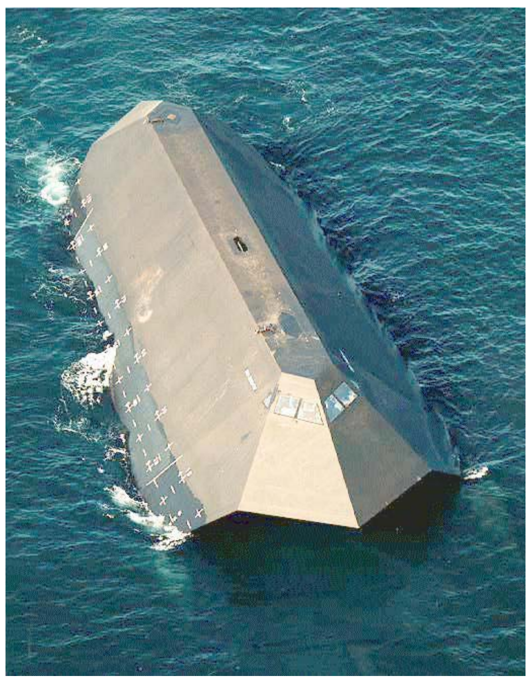
Future Combatant Concept



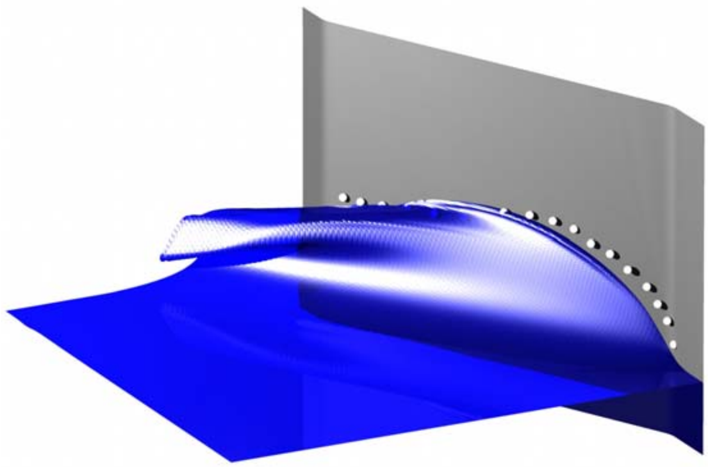


Validation of Nonlinear Free Surface Simulation for Future Combatant Concept

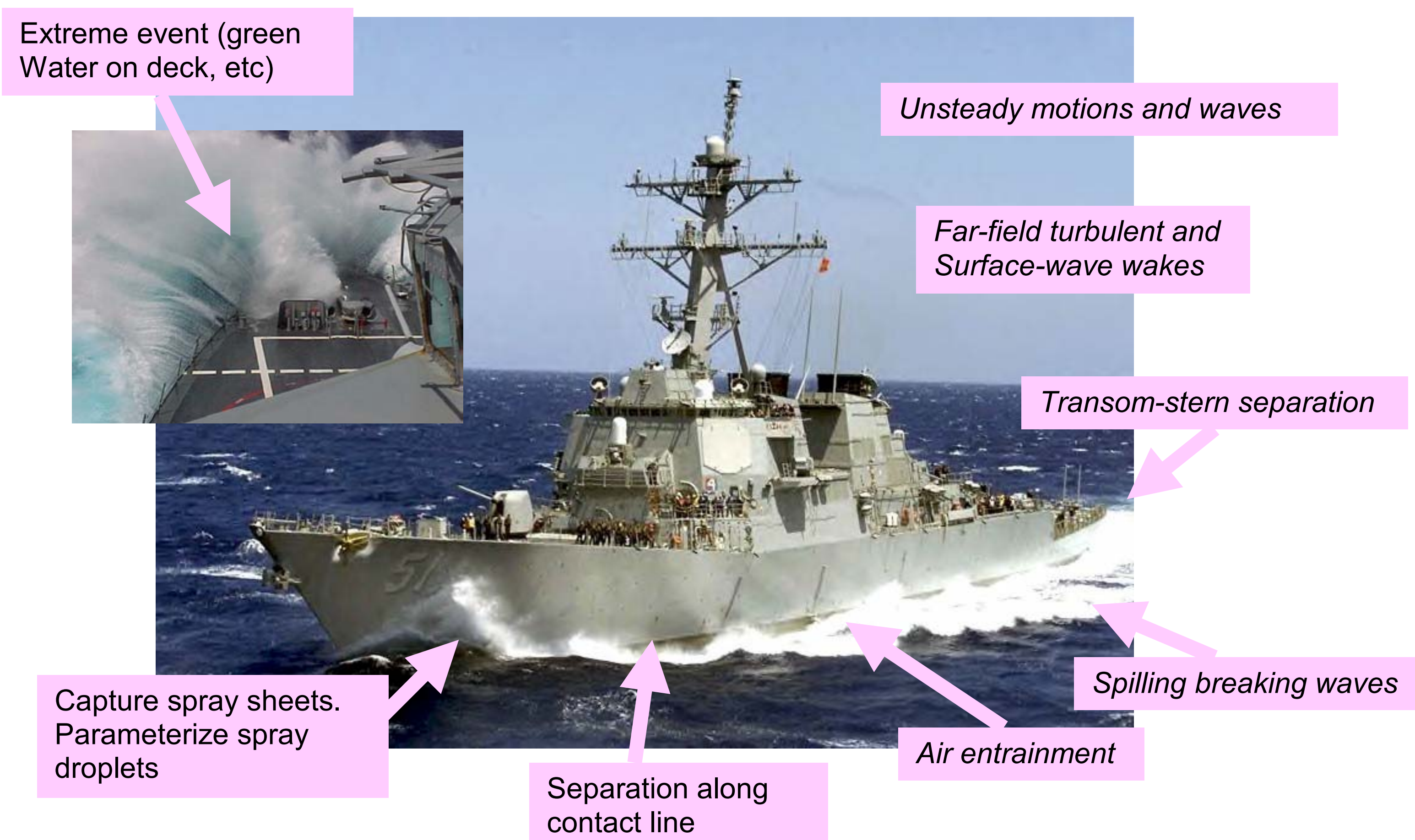




Simulation of Notional Advanced Concept Hull form with Nonlinear Free Surface



Simulations of Spray Sheet Formation near a Design Concept Hull Form



PROJECT TITLE: **STEALTHY SHIP DESIGN**

HPC COMPUTATIONAL REQUIREMENTS

HPC System	Major Software Application	Number of CPU Hours (or sustained Tera op/s) Required for Each Fiscal Year					Security Classification
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012	
T3E	UNCLE	250	500	1250	2000	2500	Unclassified
	CFDSHIP-IOWA	250	500	1250	2000	2500	Unclassified
	ShipLES	250	500	1250	2000	2500	Unclassified
	ShipLES	0	250	1000	1500	2000	**Classified**
	NFA	750	1000	2000	3500	5000	Unclassified
	NFA	0	250	1000	1500	2000	**Classified**
SP3	UNCLE	150	250	1000	2500	5000	**Classified**
	ShipLES	500	750	1500	3000	5000	Unclassified
SGI Origin	UNCLE	150	250	500	2500	5000	**Classified**

PROJECT TITLE: **STEALTHY SHIP DESIGN**

HPC COMPUTATIONAL REQUIREMENTS

HPC System	Major Software Application	Typical Memory (GB total per run) Needed Each Fiscal Year				
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012
T3E	UNCLE	32	64	128	256	512
	CFDSHIP-Iowa	32	64	128	256	512
	ShipLES	32	64	128	256	512
	ShipLES	0	32	64	128	256
	NFA	64	128	256	512	1024
	NFA	0	128	256	512	1024
SP3	UNCLE	32	64	128	256	512
	ShipLes	16	32	64	128	256
SGI Origin	UNCLE	16	32	64	128	256

PROJECT TITLE: **STEALTHY SHIP DESIGN**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW)

Major So ftware Application	Computational Technology Area	Short Code Description	Code Size *(Lines of code)*	Rate of Change *(e.g., Fraction of code that changes each year)*	Languages Used *(Fraction of each)*
UNCLE	CFD	RANS, interface tracking, level-set, structured, unstructured	100000	25%	MPI & Fortran 90
CFDSHIP-Iowa	CFD	RANS, interface tracking, level-set, structured, Chimera	50000	25%	MPI & Fortran 90
SHIPLES	CFD	LES, Level-set, structured	25000	50%	MPI & Fortran 77
NFA	CFD	LES, VOF, structured	25000	50%	HPF & MPI
CFD++ or other Commercial code	CFD	RANS/LES structured, unstructured and Chimera			

PROJECT TITLE: **STEALTHY SHIP DESIGN**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW (CONT.))

Major Software Application	Programming Models Used	Performance Critical Computation *(e.g., A linear solver or a sorting routine)*	Time to Solution Requirements *(S/W development time line requirements)*	Time to Solution Requirements *(Problem set-up time line requirements)*	Time to Solution Requirements *(Wall clock execution time for specified problem)*	Time to Solution Requirements *(Post run analysis time line requirements)*	Special Characteristics of Code *(e.g., Is it CPU performance, memory access, or I/O bound)*
UNCLE	Message Passing	Linear Solver	24 hrs	4hrs	4hrs	4hrs	CPU Performance
CFDSHIP-IOWA	Message Passing	Poisson Solver	24 hrs	4hrs	4hrs	4hrs	CPU Performance
SHIPLES	Message Passing	Poisson Solver	24 hrs	8hrs	24hrs	8hrs	CPU Performance
NFA	Message Passing and data parallel	Poisson Solver	24 hrs	8hrs	24hrs	8hrs	CPU Performance
CFD++/other				4hrs	24hrs	8hrs	CPU Performance

PROJECT TITLE: **STEALTHY SHIP DESIGN**

HPC HARDWARE REQUIREMENTS (CURRENT SYSTEM BASELINE)

HPC System	Number of Computational Nodes Used by Typical Application	Number of Processors Per Node	Processor Characteristics *(e.g., Type, nominal speed)*	Memory Per Processor	I/O System Characteristics	Total Secondary Storage	Interconnect Characteristics *(Local and aggregate bandwidth, latency)*
T3E	256	1	1.35 Gflops	512MB	500 MB/s	1 TB	650 mb/s, 2-4 ms
SP3	64	4	1.50 Gflops	512 MB	500 MB/s	1 TB	300 mb/s, 2 ms
SGI Origin	64						

PROJECT TITLE: **STEALTHY SHIP DESIGN**

HPC SOFTWARE REQUIREMENTS (APPLICATION SCALABILITY)

Major Software Application	Scalability of Code for Fixed Problem Size	Scalability of Code for Increasing Problem Size *(Fixed Workload per Processor)*	Efficiency of Code on Target Platform
UNCLE	Linear	Linear	90%
CFDSHIP-IOWA	Linear	Linear	
SHIPLES	Linear	Linear	
NFA	Linear	Linear	

PROJECT TITLE: **STEALTHY SHIP DESIGN**

HPC SOFTWARE REQUIREMENTS (DETAIL)

Major Software Application	**Memory access pattern** *(e.g., Stride 1 or random)*	**Computation profile** *(e.g., Volume of integer v. floating point computation)*	**I/O requirements** *(e.g., Volume and rate)*	**Communication characteristics** *(e.g., Describe message size and rate)*	**Inherent concurrency** *(e.g., Number of independent threads in main body of calculation)*
UNCLE	varies	Floating point: 99%	5 GB generated/day	average message size: 3 mb, 3 messages/sec	1
CFDSHIP-IOWA	varies	Floating point: 99%			
SHIPLES	1	Floating point: 99%	10 GB, 1 GB/s		
NFA	1	Floating point: 99%	10 GB, 1 GB/s		

C-5 NUCLEAR WEAPONS STOCKPILE STEWARDSHIP

Project Description

The Nuclear Weapons Stockpile Stewardship Program oversees the certification of the nuclear stockpile. In the absence of nuclear testing, stockpile certification will be based on expert analysis using validated computational simulations of the performance of nuclear weapons together with a strong above ground experimental program. The high performance computing portion of the project is the development and validation of high quality computational physics simulations of nuclear weapons and the use of those simulations for the analysis of nuclear weapons systems in support of the stockpile certification.

What is the real life problem?

The U.S. nuclear deterrent is based on the present stockpile of nuclear weapons. The designed lifetime of these weapons is about 20 years. The present stockpile systems were designed and manufactured between 1970 and 1990. With no new weapons, the U.S. will need to extend the life of the present systems for 40 to 60 or more years. The weapon systems are aging and will require maintenance and refurbishment. Problems with the weapons systems must be identified and fixes developed and implemented. An additional complication is that, due to technological progress and heightened environmental concerns, many of the processes used to manufacture the present stockpile weapons are no longer feasible. Without underground nuclear tests, the computer simulations must have vastly improved predictive capability based to the greatest extent possible on accurate physical data. The development and validation of that vastly improved predictive capability and the resulting confidence in the stockpile is the goal of the Accelerated Strategic Computing Initiative (ASCI) program.

What is the National Security Impact of (solving and not solving) the problem?

Without that confidence, the U.S. military strength will be substantially reduced. The U.S. will then be much more vulnerable to hostile powers and the security of the U.S. will be substantially reduced. In particular, we will be more vulnerable to foreign powers that possess nuclear weapons. Achieving the desired increase in predictive capability will require substantial improvements in spatial, energy, and temporal resolution, better mathematical algorithms, and more accurate physical data. Achieving the required improvements will require an increase in computer capability of roughly 10^5. High performance computing is the only way to achieve that required increase.

Why is HPC needed to address the problem?

Nuclear Weapons Stockpile Stewardship has two main technical drivers: the nuclear weapons package and the engineering components required to operate the nuclear package. Both of these efforts have the responsibility to ensure that the weapon will operate as required when it is needed and that it will not operate in any other situation.

The nuclear weapons codes being developed as part of ASCI are complex multi-physics codes. The codes integrate initial value partial differential equations for the conservation of particles, momentum, and energy for the important element and constituents of nuclear weapons.

Typical calculations use 10,000 to 1,000,000,000 mesh cells, depending on the problem and desired resolution. The larger problems must be domain decomposed to fit on available memory sizes for distributed memory systems. The partial differential equations are solved with a combination of explicit, implicit, and Monte Carlo techniques. Linear and non-liner solvers play an important role. The problems are integrated in time from an initial state to the final state. Operator splitting with some time centering handles many of the time-dependent multi-physics elements. Materials data plays an important role in determining the accuracy of the calculations.

The engineering codes are also highly complex and examine electrical and mechanical response for very complex devices subjected to a wide variety of environments. Several codes run on a common framework to allow coupled effects to be analyzed over the same engineering designs. These codes depend on solving unstructured finite element models (FEM) for a variety of mechanical, thermal, shock hydrodynamics, and crash dynamics situations. Individually, these codes do not drive the technical requirements for HPC. However, the coupled effects of such complex systems require extensive parametric studies and a need for rapid turnaround of complex simulations. These codes typically require many hundreds of processors and gigabytes of memory that is locally addressable by each node resulting in terabytes per platform.

What do you need to accomplish with HPC between 2002 and 2012?

Between 2002 and 2012, we first need to develop and validate the vastly improved computer simulation capability required for certification in the absence of testing. Then we need to apply the simulation capability to the analysis of the nuclear weapons stockpile to support certification of the stockpile. Code validation requires many parametric runs against experimental data to ensure the accuracy of the models before extrapolation to new regimes.

What do you need (H/W & S/W) from the HPC Community between 2002 and 2012 to solve your problem?

Nuclear weapons codes require thousands of processors based on the resolution and complexity of the nuclear package. The most recent calculation took 140 processor days to complete. NNSA needs a ten-fold further increase in resolution coupled with a 10- 100-fold increase in computational capability as higher-fidelity physics is added to simulations. Time to solution must be reduced from 140 days to 7 days in order to allow sufficient parametric analysis. Memory is tremendous with GBs of memory required with every processor.

As these codes run over many thousands of processors, huge data dumps must be frequently made to allow for restarts. Typical dumps require a few GBs of data to be stored every 1–2 hours. This leads to terabytes of secondary storage available with very broad I/O pathways to prevent delays in calculations.

By 2012, the computational simulations necessary for nuclear weapons stockpile certification will require a computing capability of more than 2000 teraflops per year with 200 TB of memory and 14,000 TB of secondary storage. The HPC community will need to provide the operating systems, visualization and programming environment, and code development tools to facilitate development and maintenance of the nuclear weapons computational simulations. Particularly helpful will be tools that allow faster and more effective code development and problem analysis and archiving of the results.

What could you accomplish if computers were 10x more powerful?

If the computers were 10 times more powerful (speed, memory and storage), we would be able to employ higher resolution, use more accurate methods, and run more problems. Specifically, we would probably carry out more 3-D runs using somewhat higher resolution than now planned. There are several solution methods that we would be able to employ that are substantially more accurate but are not practical with the presently planned capability. Many of our calculations are under-resolved, and more powerful computers would allow us to run problems with better convergence.

What could you accomplish if computers were 10x cheaper?

If computers were 10x cheaper, we would purchase more platforms and be able to run many more problems, especially high-resolution 2-D runs and parameter studies with 3-D runs. This would not allow more questions to be addressed but would allow us to make much better estimates of the uncertainties.

What could you accomplish if computers were 10x easier to program?

If computers were 10x easier to program, the time required to produce codes would be reduced. Application of standard software estimation methodologies validated and calibrated with metrics from the ASCI program indicates that the development of multi-physics nuclear weapons simulations should take between 7 and 9 years and will require a team of more than 20 staff. Part of the time and staff are due to the challenge of programming complex multi-physics codes for unstable and new massively parallel computers. Improving the ease of programming by a factor of 10 would reduce the time required to develop the codes to ~ 6 years. This estimate was developed by assuming, based on the experience of the present ASCI codes, that the code development tasks have the following breakdown:

10% requirements gathering

10% design and planning

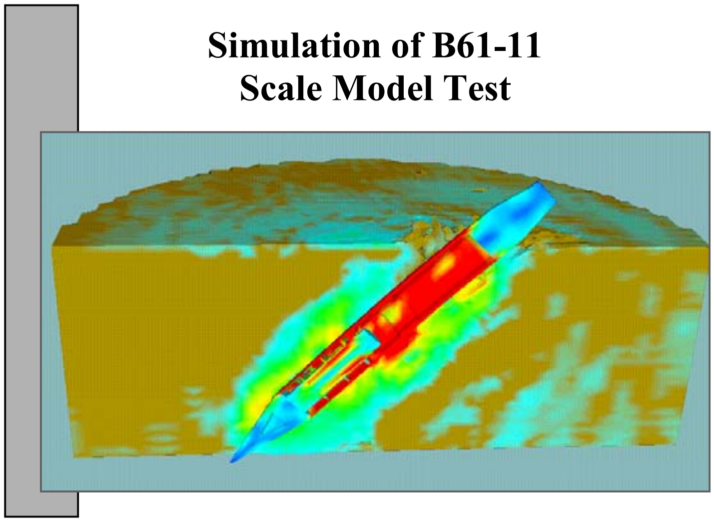
30% algorithm development

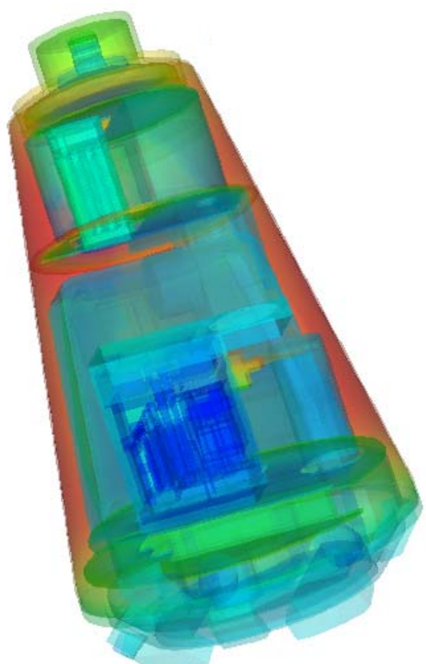
25% parallel programming

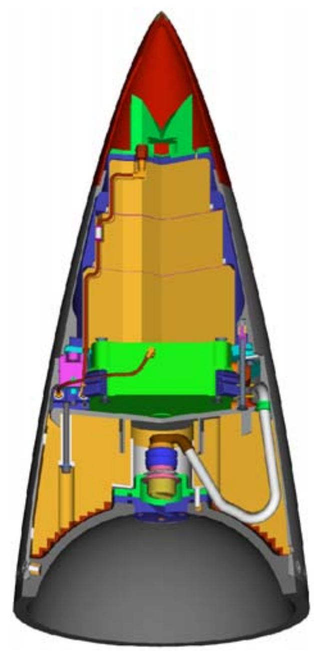
10% documentation

15% testing.

We would be able to reduce the parallel computing from 25% to 2.5% and perhaps reduce the algorithm development from 30% to 25%. This is a roughly 30% reduction in the total code development yielding roughly a code development time of 6 years from 8 years.


Simulation of B61-11
Scale Model Test





PROJECT NAME: **NUCLEAR WEAPONS STOCKPILE STEWARDSHIP**

HPC COMPUTATIONAL REQUIREMENTS

HPC System	Major Software Application	Number of CPU Hours (or sustained Tera op/s) Required Each Fiscal Year					Security Classification
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012	
ASCI Red (Sandia Nat Lab)	Nuclear Weapons Engineering Design and Certification	2.7 T	—	—	—	—	Classified
ASCI Blue Mountain (Los Alamos)	Nuclear Weapons Certification (3 code projects)	3T	—	—	—	—	Classified
ASCI White (Livermore Nat. Lab.)	Nuclear Weapons Certification (3 code projects)	3T	3T	—	—	—	Classified
ASCI White Equiv (Livermore Nat. Lab.)	Nuclear Weapons Certification (2 code projects)	2.1T	8.1T	56T	175T	872T	Classified
ASCI White Equiv (Livermore Nat. Lab.)	Nuclear Weapons Materials (2 code projects)	0.75T	3T	19T	60T	299T	Classified
ASCI White (Livermore Nat. Lab.)	Nuclear Weapons Engineering Design and Certification	3 T	3 T	—	—	—	Classified

PROJECT NAME: **NUCLEAR WEAPONS STOCKPILE STEWARDSHIP**

HPC COMPUTATIONAL REQUIREMENTS

HPC System	Major Software Application	Number of CPU Hours (or sustained Tera op/s) Required Each Fiscal Year					Security Classification
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012	
ASCI Q (Los Alamos)	Nuclear Weapons Certification (3 code projects)	10T	10T	10T	—	—	Classified
ASCI Q (Los Alamos)	Nuclear Weapons Engineering Design and Certification	5 T	8 T	8 T	8 T	—	Classified
ACSI Future Systems	Nuclear Weapons Certification (3 code projects)	45 T	221 T	322 T	592 T	820 T	Classified
All ASCI systems*	Nuclear Weapons Certification	74.6 T	256 T	415 T	835 T	1991 T	Classified

* Requested, would be used for certification and code development if available

PROJECT NAME: **NUCLEAR WEAPONS STOCKPILE STEWARDSHIP**

HPC MEMORY REQUIREMENTS

HPC System	Major Software Application	Typical Memory (GB total per run) Needed Each Fiscal Year				
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012
ASCI White (Livermore Nat. Lab.)	Nuclear Weapons Certification (3 code projects)	1 TB	3 TB	4 TB	6 TB	10 TB
ASCI White (Livermore Nat. Lab.)	Nuclear Weapons Certification (2 code projects)	1.3 TB	2.2 TB	3.5 TB	6 TB	9 TB
ASCI White (Livermore Nat. Lab.)	Nuclear Weapons Materials (2 code projects)	850 GB	1.4 TB	2.1 TB	3.5 TB	5.5 TB
ASCI Red (Sandia Nat Lab)	Nuclear Weapons Engineering Design and Certification (5 Code Projects)	2.3 TB	5.3 TB	5.7 TB	7.4 TB	8.9 TB
All ASCI systems*	Nuclear Weapons (12 code projects)	5.5 TB	11.9 TB	15.3 TB	22.9 TB	33.4

* Requested, would be used for certification and code development if available

PROJECT NAME: **NUCLEAR WEAPONS STOCKPILE STEWARDSHIP**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW)

Major Software Application	Computational Technology Area	Short Code Description	Code Size *(Lines of code)*	Rate of Change *(e.g., Fraction of code that changes each year)*	Languages used *(Fraction of each)*
Nuclear Weapons Certification (3 code projects)	NWS	initial value, implicit and explicit partial differential equations for 2D and 3D density, momentum and energy conservation, 10^4 to 10^9 cells, multi-physics CFD, rad-hydro, transport	3 codes, 300,000 to 800,000 lines for each code	up to 150,000 lines per year	C: 4%, C++: 20%, F90:76%
Nuclear Weapons Certification (2 code projects)	NWS	initial value, implicit and explicit partial differential equations for 2D and 3D density, momentum and energy conservation, 10^4 to 10^9 cells, multi-physics CFD, rad-hydro, transport	2 Codes, 400,000 to 1,000,000 lines for each code	10-20% of Code, all modules touched	C: 3 or 90%, C++: 80 or 5%, F: 7 or 5%, Python: 10 or 0%
Nuclear Weapons Materials	CFD	Turbulance simulation	250,000 lines	10-15% of code	C: 10%, C++: 80%, F:10%
Nuclear Weapons Materials	CCM	1st Principles material models	250,000	20-25% if code	C: 40%, C++: 50%, F: 10%

PROJECT NAME: **NUCLEAR WEAPONS STOCKPILE STEWARDSHIP**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW)

Major Software Application	**Computational Technology Area**	**Short Code Description**	**Code Size** *(Lines of code)*	**Rate of Change** *(e.g., Fraction of code that changes each year)*	**Languages used** *(Fraction of each)*
Alegra/Nevada	CSM/CFD/CEA	Large-deformation mechanics for multiphysics shock hydrodynamics, both FEM (unstructured) and FV (structured) modes	1000k	10% -20%	95% C++; 5% C and Fortran
Major Software Application	Computational Technology Area	Short Code Description	Code Size (lines of code)	Rate of Change (e.g., fraction of code that changes each year)	Languages used (fraction of each)
Salinas	CSM	Unstructured FEM structural mechanics code for modal analysis	600k	Same	95% C++; 5% C and Fortran
Presto	CSM	Unstructured FEM non-linear mechanics code for crash dynamics	600k	Same	95% C++; 5% C and Fortran
ITS	CRT	MC radiation transport code for thermal-mechanical loading	200k	Same	Fortran

PROJECT NAME: **NUCLEAR WEAPONS STOCKPILE STEWARDSHIP**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW)

Major Software Application	Computational Technology Area	Short Code Description	Code Size *(Lines of code)*	Rate of Change *(e.g., Fraction of code that changes each year)*	Languages used *(Fraction of each)*
Calore	CTM	Unstructured FEM thermal-chemical analysis code for heat transfer	600k	Same	95% C++; 5% C and Fortran
			Key (for Computational Technology Area types)		
			Computational Structural Mechanics (CSM)		
			Computational Fluid Dynamics (CFD)		
			Computational Electromagnetics and Acoustics (CEA)		
			Nuclear Weapons Simulations (NWS)		
			Computational Chemistry and Materials Science (CCM)		

PROJECT NAME: **NUCLEAR WEAPONS STOCKPILE STEWARDSHIP**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW (CONT.))

Major Software Application	Programming Models Used	Performance Critical Computation *(e.g., A linear solver or a sorting routine)*	Time to Solution Requirements *(S/W Development time line requirements)*	Time to Solution Requirements *(Problem set-up time line requirements)*	Time to solution requirements *(Wall clock execution time for specified problem)*	Time to Solution Requirements *(Post run analysis time line requirements)*	Special Characteristics of Code *(e.g., Is it CPU performance, memory access, or I/O bound)*
Nuclear Weapons Certification (5 code projects)	MPI, Open MP	linear solvers, Monte Carlo	Full code—5 to 10 years	2D—one day to several weeks; 3D—one week to several months	2D—one day to several weeks; 3D—one week to several months	2D—one day to several dayss;;3D—one week to several weeks	memory access
Nuclear Weapons Materials (2 code projects)	MPI, Open MP	FFT, nonlinear equation solvers	Full code—5 to 10 years	2D—one day to several weeks; 3D—one week to several months	2D—one day to several weeks;;3D—one week to several months	2D—one day to several days; 3D—one week to several weeks	memory access, communication bound
Alegra	MPI	stress divergence, contact algorithms, linear solver	~10 years for full code, 2-5 yrs. for major capability	~ 1 Mo.	~ 1 week	~1 week - 1 month	1) memory access, 2) network bandwidth
Salinas	MPI	eigen solver, dynamic load balancing	Same	Same	Same	Same	1) memory access, 2) network bandwidth

PROJECT NAME: **NUCLEAR WEAPONS STOCKPILE STEWARDSHIP**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW (CONT.))

Major Software Application	Programming Models Used	Performance Critical Computation *(e.g., A linear solver or a sorting routine)*	Time to Solution Requirements *(S/W Development time line requirements)*	Time to Solution Requirements *(Problem set-up time line requirements)*	Time to solution requirements *(Wall clock execution time for specified problem)*	Time to Solution Requirements *(Post run analysis time line requirements)*	Special Characteristics of Code *(e.g., Is it CPU performance, memory access, or I/O bound)*
Presto	MPI	contact algorithms	Same	Same	Same	Same	1) memory access, 2) network bandwidth
Major Software Application	Programming models used	Performance critical computation (e.g., a linear solver or a sorting routine)	Time to solution requirements (S/W development time line requirements)	Time to solution requirements (problem set-up time line requirements)	Time to solution requirements (wall clock execution time for specified problem)	Time to solution Requirements (Post run analysis time line requirements)	Special characteristics or code (e.g., is it CPU performance, memory access, or I/O bound)
ITS	MPI	ray tracing	Same	Same	Same	Same	1) memory access, 2) network bandwidth
Calore	MPI	radiation view factor and chemistry	Same	Same	Same	Same	1) memory access, 2) network bandwidth

PROJECT NAME: **NUCLEAR WEAPONS STOCKPILE STEWARDSHIP**

HPC HARDWARE REQUIREMENTS (CURRENT SYSTEM BASELINE)

HPC System	Number of Computational Nodes Used by Typical Application	Number of Processors Per Node	Processor Characteristics *(e.g., Type, nominal speed)*	Memory Per Processor	I/O System Characteristics	Total Secondary Storage	Interconnect Characteristics *(Local and aggregate bandwidth, latency)*
ASCI Blue Mountain (Los Alamos)	10	128	MIPS RS10K (250MHz)	0.25 GB	parallel global file system	76 TB	HIPPI 600MB/s local (7μsec) 48GB/s aggregate (15μsec)
ASCI Red (SNL)	1000	2 (both can be usefully employed in some computations, in others only 1)	Pentium II @ 330 MHz	.25 GB	parallel global file system @ ~150MB/s	12.5TB	Custom backplane with 800 Mbytes/sec node link bandwidth (bi-directional), 51.2 Gbytes/s cross-sectional bisection bandwidth (bi-directional), 15 microsecond latency
ASCI Blue Pacific (LLNL)	5100 total	4	Power PC 604e (332MHz)	0.375, 0.625	parallel global file system	62.5 TB	TB3/HPGN
ASCI White (LLNL)	128—192	16	Power 3 (375 MHz)	1 GB	parallel global file system	145 TB	Colony DS

PROJECT NAME: **NUCLEAR WEAPONS STOCKPILE STEWARDSHIP**

HPC HARDWARE REQUIREMENTS (CURRENT SYSTEM BASELINE)

HPC System	Number of Computational Nodes Used by Typical Application	Number of Processors Per Node	Processor Characteristics *(e.g., Type, nominal speed)*	Memory Per Processor	I/O System Characteristics	Total Secondary Storage	Interconnect Characteristics *(Local and aggregate bandwidth, latency)*
ASCI Q (Los Alamos)	500—1000 with a few of 2000—3000	4	Ev68 (1.25GHz)	2—8 GB	parallel global file system	700 TB	Quadrics 700MB/s local (5μsec) 750GB/s aggregate (10μsec)
HPC System	Number of computational nodes used by typical application	Number of processors per node	Processor characteristics (e.g., type, nominal speed)	Memory per processor	I/O system characteristics	Total Secondary Storage	Interconnect characteristics (local and aggregate bandwidth, latency)
ASCI Purple (LLNL)			60TeraOPS peak	30TB total		1200TB	
ASCI Red Storm (SNL)			20 TeraOPS peak				

PROJECT NAME: **NUCLEAR WEAPONS STOCKPILE STEWARDSHIP**

HPC SOFTWARE REQUIREMENTS (APPLICATION SCALABILITY)

Major Software Application	Scalability of Code for Fixed Problem Size	Scalability of Code for Increasing Problem Size *(Fixed Workload per Processor)*	Efficiency of Code on Target Platform
Nuclear Weapons Certification (5 code projects)	low priority, small problems run on small node counts (less than 4 processors), larger problems scale to 4000 processor or more	Targeted standard operating procedure, fill memory, up to 6000 processors	parallel efficiency: 30 to 70% for large problems, up to 90% for small problems; processor efficiency—2-20% irregular data structures (memory bottlenecks)
Nuclear Weapons Materials (2 code projects)	low priority, small problems run on small node counts (less than 4 processors), larger problems scale to 4000 processor or more	Targeted standard operating procedure, fill memory, up to 6000 processors	parallel efficiency: 30 to 70% for large problems, up to 90% for small problems; processor efficiency—2-20% irregular data structures (memory bottlenecks)
Alegra	problem dependent but typically very good out to 200 processors	very good (80%-90%) out to 2000+ processors	~10% unstructured, 20% structured
Salinas	Same	Same	~20% of peak
Presto	Same	Same	~10% of peak
ITS	Same	Same	10% of peak
Calore	Same	Same	10% of peak

PROJECT NAME: **NUCLEAR WEAPONS STOCKPILE STEWARDSHIP**

HPC SOFTWARE REQUIREMENTS (DETAIL)

Major Software Application	**Memory Access Pattern** *(e.g., Stride 1 or random)*	**Computation Profile** *(e.g., Volume of integer v. floating point computation)*	**I/O Requirements** *(e.g., Volume and rate)*	**Communication Characteristics** *(e.g., Describe message size and rate)*	**Inherent Concurrency** *(e.g., Number of independent threads in main body of calculation)*
Nuclear Weapons Certification (3 code projects)	irregular, but not entirely random, stride 1, working arrays, domain decomposition, load distribution	99% floating point computation except for Monte Carlo (50% floating point, 50% integer)	100 MB-500GB/hour for restart files in bursts; greater than 1 GB/s to parallel shared file systems;	generally small, less than 10kB/message, some codes have ~ 1 MB/message; lots of messages	up to millions of concurrent operations for 2D spatial grids, multi-millions of concurrent operations in 3D spatial grids
Nuclear Weapons Certification (2 code projects)	random and hybrid random with structured access	Mostly floating point	2-10 GB per write, a write every 2 hours	varies, small bursty to infrequent very large	varies from 1-16
Nuclear Weapons Materials (2 code projects)	hybrid random with structured access and nearest neighbor with global reduction	one integer, one floating point	1-4 GB per write every hour	heavy communication in small messages	varies from 1-16
Major Software Application	Memory access pattern (e.g., stride 1 or random)	Computation profile (e.g., volume of integer v. floating point computation)	I/O requirements (e.g., volume and rate)	Communication characteristics (e.g., describe message size and rate)	Inherent concurrency (e.g., number of independent threads in main body of calculation)

PROJECT NAME: **NUCLEAR WEAPONS STOCKPILE STEWARDSHIP**

HPC SOFTWARE REQUIREMENTS (DETAIL)

Major Software Application	**Memory Access Pattern** *(e.g., Stride 1 or random)*	**Computation Profile** *(e.g., Volume of integer v. floating point computation)*	**I/O Requirements** *(e.g., Volume and rate)*	**Communication Characteristics** *(e.g., Describe message size and rate)*	**Inherent Concurrency** *(e.g., Number of independent threads in main body of calculation)*
ALEGRA (unstructured mode)	cache-based, random within cache	65% floating point	very fast dump of full memory to disk every 4 hours for checkpoint/restart	communications intensive code with variable size messages, e.g. 10Kbyte - 100Kbyte messages @ 3GBytes/sec	currently all MIMD codes that scale to thousands of processors; single-threaded on each processor - more threading perhaps possible
ALEGRA (structured mode)	stride 1	85% floating point	Same	Same	Same
Salinas	cache-based, random within cache	very predominantly floating point	Same	Same	Same
Presto	cache-based, random within cache	very predominantly floating point	Same	Same	Same
ITS	cache-based, random within cache	very predominantly floating point	Same	Same	Same
Calore	cache-based, random within cache	very predominantly floating point	Same	Same	Same

C-6 SIGNAL AND IMAGE PROCESSING

Project Information

HPC Project Name: Signals and Image Processing

Agency and Organization: Naval Reconnaissance Office – Imagery Intelligence

Project Leader(s): Robert Alexander/Joe Swartz/Hank Dardy (NRL)

Project Leader E-Mail Address: rja@mindspring.com

Project Leader Phone Number: 703-808-1993

Project Leader Address: IMINT/RTS, 23G06N, Chantilly, Virginia 20151

Number of Users in Project: Several small groups

Computational Technology Area(s) (See attached list): Image processing for Intelligence and Earth Sensing Applications

Project Description

Image processing – to include radiometric, multi-spectral, and hyperspectral data – of tactical and National assets. Also includes potential processing of synthetic aperture data for both passive and active radar and electro-optical data. Concept of operations encompasses real-time collection at ground stations, automated processing, archiving of raw data, and warehousing of processed data. Mission partners (NIMA, Central MASINT office, etc.) are responsible for dissemination of products via library servers.

Other areas of future concern and hence, of R&D interest now, are the distribution or re-routing of processing loads, making software portable across heterogeneous high performance platforms, making software operable on commodity clusters, streamlining the expensive and time-consuming path required for developing an algorithm from engineering code to high performance software, and insertion of new software into legacy systems. Immersive, remote visualization for real-time situational awareness presentations of processed data to Seniors or CINCs is also of major importance.

What is the real life problem?

Real-time image (and signal) reconstruction, processing, exploitation, visualization, and dissemination for use in area search, site reconnaissance, mission planning, tactical situation awareness, targeting, battle damage assessment, and supporting challenging and complex intelligence problems as required.

What is the National Security Impact of (solving and not solving) the problem?

Significant-to-critical, with increasing criticality over time. *"Not solving"* is not *"an acceptable answer!"*

Why is HPC needed to address the problem?

HPC is necessary due to the complexity of processing large amounts of data in time to exploit information gained while it is still of use to the warfighter. Real-time tasking and exploitation of sensory systems is critical for situation awareness.

What do you need to accomplish with HPC between 2002 and 2012?

Improvements of several orders of magnitude in processing power with well balanced architecture capable of handling/processing/disseminating streams of 'big data' capable of moving large, continuous flows both within architected processing systems and continuous real-time ingress from sensors and platforms, and egress to projectors/displays of data for knowledge-based situation awareness.

What do you need (H/W & S/W) from the HPC Community between 2002 and 2012 to solve your problem?

Better compilers that leverage optimizing the underlying hardware, optimized software math libraries, ability to handle continuous high data rate streams (input and output), shared memory without excessive latency penalty, multi-teraflops, improved integer processing, and better packaging density (with lower power consumption per flop).

Project Vision

What is the mission vision?

Handling today's intractable problems so as to process more data at higher resolution in less time in support of a greater number of threats. Improve data gathering and collection systems processing for future space and ground sensor systems. Processing system needs to be flexible to accommodate tasking with priority and preemption. Needs to merge conventional Von Neumann processing with data flow constructs to create a hybrid supercomputer architecture.

What could you accomplish if computers were 10x more powerful?

Reduce system latency and process more data immediately; increase sensor resolution, area coverage, and communication rates to greatly affect time-to-solution. Enhance quality and quantity of finished product for end decision makers who may need to handle multiple threats. *Note: The requirement for a balanced architecture remains – more powerful systems must remained balanced. Increased processor performance yields little improvement if I/O bandwidth doesn't improve as well. System has to expose interface sufficiently to end-user so that reasonably astute programmers can leverage this power easily and effectively.*

What could you accomplish if computers were 10x cheaper?

10x cheaper computers would eliminate the need for stove-piped or system-specific architectures. 10x cheaper computers would also make it easier to provide added value processing so that end users could get a more finished product without having the need to perform specialized processing on their own. Commensurate communications and networking capabilities are also needed.

What could you accomplish if computers were 10x easier to program?

10x easier programming would make it possible to field more new algorithms and to accomplish that more quickly. Insertion of new algorithms into existing architecture is probably the most significant impedance to high performance computing today.

10x easier programming would make it easier for analysts of National and tactical data to pose and purse hypotheses more easily across multiple sources of data (i.e., multi-INT). 10x processing would better facilitate the application of information technology to the exploitation of National and tactical data.

PROJECT NAME: **SIGNALS AND IMAGE PROCESSING**							
HPC COMPUTATIONAL REQUIREMENTS							
HPC System	**Major Software Application**	**Number of CPU Hours (or sustained Tera op/s) Required for Each Fiscal Year**					**Security Classification**
		FY 2003	**FY 2005**	**FY 2007**	**FY 2010**	**FY 2012**	
A	2d-FFTs	100G	500GF	2TF	5TF	10TF	TS
B	Signals/Image Processing	2,600GF	4,400GF	8TF	15TF	24TF	TS

PROJECT NAME: **SIGNALS AND IMAGE PROCESSING**

HPC MEMORY REQUIREMENTS

HPC System	Major Software Application	Typical Memory (GB total per run) Needed Each Fiscal Year				
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012
A	2d-FFT	16-32GB	64-128GB	128-256GB	1.0TB	2.0TB+
B	Signals/Image Processing	1,500GB	3,000GB	5,500GB	10.0TB	2.4TB+

PROJECT NAME: **SIGNALS AND IMAGE PROCESSING**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW)

Major Software Application	Computational Technology Area	Short Code Description	Code Size *(Lines of code)*	Rate of Change *(e.g., Fraction of code that changes each year)*	Languages Used *(Fraction of each)*
1d-FFT	SIP		~1000	~10%	Fortran & C
2d-FFT	SIP		~1000	~12%	Fortran & C
Vector mul	SIP		~1000	~10%	Fortran & C
Special macros	SIP		~2000	~5%	Assembly
				Key (for Computational Technology Area types)	
				Signal Image Processing (SIP)	

PROJECT NAME: **SIGNALS AND IMAGE PROCESSING**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW (CONT.))

Major Software Application	Programming Models Used	Performance Critical Computation *(e.g., A linear solver or a sorting routine)*	Time to Solution Requirements *(S/W development time line requirements)*	Time to Solution Requirements *(Problem set-up time line requirements)*	Time to Solution Requirements *(Wall clock execution time for specified problem)*	Time to Solution Requirements *(Post run analysis time line requirements)*	Special Characteristics of Code *(e.g., Is it CPU performance, memory access, or I/O bound)*
2d-FFTs	SHMEM	butterfly, corner turning, polar-rect conv.	xx	xx	real-time	real-time	all of the above
2d-FFTs	MPI, OpenMPI	butterfly, corner turning, polar-rect conv., integer performance	xx	xx	real-time	real-time	all of the above

PROJECT NAME: **SIGNALS AND IMAGE PROCESSING**

HPC HARDWARE REQUIREMENTS (CURRENT SYSTEM BASELINE)

HPC System	Number of Computational Nodes Used by Typical Application	Number of Processors Per Node	Processor Characteristics *(e.g., Type, nominal speed)*	Memory Per Processor	I/O System Characteristics	Total Secondary Storage	Interconnect Characteristics *(Local and aggregate bandwidth, latency)*
A	128	1	1.0GF/proc	1 GB	24GBps	20TB	400MB mBW, 250-350 ns rts
B	768	1	0.5GF/proc	1 GB	1.6GBps	150TB	400MB mBW, 250-350 ns rts
Sig/Img Proc	2,600	1	1.0GF/proc	1 GB	3.5GBps	1PB+	400MB mBW, 250-350 ns rts

PROJECT NAME: **SIGNALS AND IMAGE PROCESSING**

HPC SOFTWARE REQUIREMENTS (APPLICATION SCALABILITY)

Major Software Application	Scalability of Code for Fixed Problem Size	Scalability of Code for Increasing Problem Size *(Fixed Workload per Processor)*	Efficiency of code on target platform
2d-FFTs	Scalable	Scalable	15%

PROJECT NAME: **SIGNALS AND IMAGE PROCESSING**

HPC SOFTWARE REQUIREMENTS (DETAIL)

Major Software Application	**Memory Access Pattern** *(e.g., Stride 1 or random)*	**Computation Profile** *(e.g., Volume of integer v. floating point computation)*	**I/O Requirements** *(e.g., Volume and rate)*	**Communication Characteristics** *(e.g., Describe message size and rate)*	**Inherent Concurrency** *(e.g., Number of independent threads in main body of calculation)*
1d-FFT	stride-1	floating	TB at 10 Gbps flows	2.5-10 Gbps+	32
2d-FFT	stride-2K	floating	TB at 10 Gbps flows	2.5-10 Gbps+	32000
cmplx mul	stride-1	floating	streams flows	2.5-10 Gbps+	1000

C-7 ARMY FUTURE COMBAT SYSTEMS

Project Information

HPC Project Name: Army Transformation: Future Combat Systems (FCS)/Army 2010+

Agency and Organization: U.S. Army Research Laboratory

Project Leader(s): a) Kent Kimsey, b) David Kleponis, and c) Raju Namburu

Project Leader E-Mail Address: a) kimsey@arl.army.mil, b) kleponis@arl.army.mil, and c) raju@arl.army.mil

Project Leader Phone Number: a) 410-278-6083, b) 410-278-6803, and c) 410-278-0274

Project Leader Address: a) Director, U.S. Army Research Laboratory, ATTN: AMSRL-WM-TC (K. Kimsey), Aberdeen Proving Ground, MD 21005-5066 and c) Director, U.S. Army Research Laboratory, ATTN: AMSRL-CI-HC (Raju Namburu), Aberdeen Proving Ground, MD 21005-5066

Number of Users in Project: 57

Computational Technology Area(s) (See list in companion spreadsheet, sheet 3): Computational Structural Mechanics (CSM) and Computational Electromagnetics and Acoustics (CEA)

Project Description

Large-scale simulations are an integral part of research programs conducted at the U.S. Army Research Laboratory (ARL). Many of these programs are focused on the development and evaluation of lethality and survivability technologies germane to future land combat systems (e.g., FCS). ARL's research programs leverage high performance computing (HPC) to solve problems associated with the discovery, refinement, development, and evaluation of critical lethality and survivability technologies. Examples include advanced kinetic energy (KE) penetrators, multi-functional warheads (blast, fragmenting, shaped charge, explosively formed penetrator, and others), kinetic energy missiles, passive armors, reactive armors, and other advanced/hybrid armors. The penetration mechanics of advanced KE penetrators, as well as novel penetrator concepts impacting modern and emerging armors, is complex. Only a portion of the dynamics of penetrator-target interaction can be gleaned solely from terminal ballistic experiments. A large portion of these dynamics must be examined with modeling and simulation. Researchers at ARL use a combined numerical and experimental approach in research studies to: 1) identify critical armor/armament defeat mechanisms; 2) characterize the performance of advanced ballistic materials such as ceramics, composites, advanced metal alloys and energetic materials; 3) explore synergistic combinations of highly efficient defeat mechanisms to develop hybrid armor technologies; and 4) evaluate the efficacy of advanced KE penetrators and novel penetrator concepts for defeating modern and emerging armor technologies. This combined approach enables researchers to advance the state-of-the-art by accelerating the tempo of research programs. Furthermore, this combined approach is critical to enabling researchers to discover and rank technical approaches so that experimental programs focus on the most promising technologies and concepts.

A better understanding of the entire spectrum of signatures for combat systems or a system of systems in a realistic battlefield environment is key in achieving dominance on the battlefield. The development and deployment of future systems requires advanced predictive design and prototyping that rely heavily on modeling and simulation to assess feasibility, optimize performance, and trade-off competing requirements. These challenges can be addressed effectively by using selected experiments, scalable computing resources, validated algorithms, scalable software, and multidisciplinary computational environments. Low observable survivability requirements addressed in this study include low/high frequency RF spectrum, IR, coatings, and acoustics in the presence of the battlefield environment (including terrain and weather effects).

What is the real life problem?

The Chief of Staff of the Army (CSA), GEN Eric K. Shinseki, has articulated a vision to transform the Army. The CSA's intent is to make heavy forces more strategically responsive and light forces more lethal and survivable. Strategic responsiveness means deploying, anywhere in the world: a brigade in 96 hours, a division in 120 hours, and five divisions in 30 days. In order to achieve these deployment timeframes, a Brigade Combat Team must fit all materiel required for deployment on a C-130 cargo aircraft. This means that each combat system or combat platform must weigh less than 20 tons and be tailored to achieve the ground combat and mobility requirements essential to battlefield dominance.

The science and technology (S&T) community plays a paramount role in the long-term transformation of the Army into the 21st Century. The Army's highest priority S&T initiative enabling this new vision is the Future Combat Systems (FCS) Program. The program goal is to create responsive, deployable, agile, versatile, lethal, survivable, and sustainable combat systems that can enter production in 2008 and be fielded as early as 2010. Maintaining or increasing lethality and survivability of current heavy combat systems in less than 1/3 the weight illustrates the need for significant technological advances and clearly defines the real life problem.

What is the National Security Impact of (solving and not solving) the problem?

Achieving the Army's transformation goals for an objective force that is strategically responsive and dominant across the full spectrum of operations will enable future Army operations; from small-scale contingencies including anti-terrorism operations to major theater wars. Advances in information, materiel, and weapons systems technologies will make it possible for objective force units to achieve at least the same effectiveness as today's forces, but with fewer, lighter, and more sustainable systems.

Why is HPC needed to address the problem?

HPC applications in computational terminal ballistics are an integral part of research programs to develop and evaluate advanced lethality and survivability technologies germane to advanced combat systems including the Army's FCS program. Large-scale simulations capture many of the dynamics of complex projectile-target interactions that cannot be gleaned solely from terminal ballistic experiments. Today, modeling emerging and advanced lethality and survivability technologies mandate higher fidelity models due to 1) larger computational domains, 2) finer mesh resolutions to model complex geometries of interest, 3) simulation times that lie in the millisecond regime as opposed to the microsecond regime, and 4) implementation of advanced and or multidisciplinary physics as well as material models. These factors, coupled with the requirement to conduct myriad three-dimensional parametric design simulations, are strong stimuli for exploiting scalable HPC systems. Furthermore, the accelerated pace required for technology development, evaluation, and demonstration to enable the Army's transformation does not permit traditional incremental advances in technology solely through experimentation. Leap-ahead advances in lethality and survivability technologies can only be achieved using a judicious combination of large-scale simulations coupled with experiments. Scalable HPC applications are a critical component of research and development programs and they play a paramount role in screening lethality and survivability technologies or concepts which are becoming more costly and remain very time consuming to evaluate using experiments only.

Future Combat Systems (FCS) and Army 2010+ are envisioned to use new materials and composite construction to meet the weight, lethality, and survivability criteria. Modeling the complex geometries of full-scale combat systems, antennas, composite layers, and construction of the system with different dielectric properties present significant gridding challenges. The capability of modern radar systems to detect, identify, and target battlefield assets is steadily advancing through the exploitation of enhanced resolution and imaging capabilities available by using the millimeter wave (MMW) region of the electromagnetic (EM) spectrum. Large-scale physics based simulations will assist in a better understanding of the full electromagnetic spectrum. However, modern radar systems dictate very fine resolution meshes to numerically capture high frequency content of the EM spectrum. Typically, grid size is inversely proportional to the frequency; that is, finer and finer resolution is needed to capture high frequency waves. For these examples, computational and memory requirements dictate the necessity of scalable computers. For example, to accurately simulate RCS at 35 GHz for a typical Army armored vehicle requires tera-cell (10E+12 finite volume cells) models. The computational grid size increases substantially not only for higher frequencies but also for simulating the same size vehicle and incorporating or adding new materials/layers, the ground plane, battlefield effects, etc.

What do you need to accomplish with HPC between 2002 and 2012?

Achieving technology readiness levels for the Army's FCS program is the focus of Army S&T initiatives during the 2002–2012 timeframe. Researchers will use a judicious combination of large-scale simulations of complex projectile-target interactions coupled with terminal ballistic experiments to enhance the overall survivability and lethality of armored ground vehicles. The lethality component will focus on energy-efficient lethal mechanisms for large caliber and medium caliber ammunition, overwhelming lethal mechanisms for kinetic energy missiles, and multifunctional warheads with enhanced blast effects. The survivability component will focus on innovative materials, structures, integrated armor concepts, protection and defeat mechanisms, and survivability concepts for lightweight protection against a spectrum of battlefield threats.

What do you need (H/W & S/W) from the HPC Community between 2002 and 2012 to solve your problem?

Large, scalable HPC systems (100 TFLOPS capability) in secure environments with balanced storage and archiving systems.

Scalable scientific visualization systems and scalable visualization SW to visualize giga-cell models.

Coupled or multidisciplinary (CSM, CFD, CEA) SW for shock physics problems with improved physics.

Development of multidisciplinary design optimization to assist signature management of combat systems.

Hybrid approaches for solving a wide spectrum of electromagnetic frequencies.

Software that couples IR, acoustic, and visual signature simulations in the presence of a battlefield environment

Project Vision

What is the mission vision?

The project vision is to conduct a judicious combination of large-scale simulations of complex projectile-target interactions coupled with terminal ballistic experiments to discover, refine, develop, and evaluate advanced lethality and survivability technologies germane to the advanced combat systems such as the Army's Future Combat Systems (FCS). The vision leverages HPC applications coupled with experiments to assure the efficacy of legacy and interim forces with the attainment of the Army's technology readiness goals for FCS and beyond, i.e., Army 2010+.

The mission also encompasses a full spectrum of HPC signature modeling capabilities to assist survivability and lethality technologies towards the Army Transformation. Through exploitation of HPC technologies and virtual electromagnetic effects, prototyping tools improve low observable technologies, optimize performance, and trade-off competing requirements for a wide class of present and future Army systems.

What could you accomplish if computers were 10x more powerful?

The Army's FCS program can be divided into three research and development phases; 1) concept and technology development, 2) systems integration and demonstration, and 3) initial production and development. To date, HPC applications have focused on discovery, refinement, development, and evaluation of advanced lethality and survivability technologies germane to the FCS. As the Army's FCS program enters the systems integration/demonstration phase in FY04, the scope of HPC applications must be broadened to extend their design utility to the evaluation and optimization of integrated survivability technologies (e.g., active protection, hybrid armors, and signature management) as well as energy efficient lethality technologies for both cannon and missile platforms. Beyond FY07, HPC applications will support the discovery, development, evaluation, and optimization of follow-on lethality and survivability technologies to meet the Army's S&T objectives for product improvement programs as well as next generation weapon systems (i.e., Army 2010+) to assure continued battlefield dominance. The primary S&T benefits attainable via the acquisition of HPC systems with a tenfold increase in capability are:

A broadened scope of current HPC parametric analyses to encompass complex projectile target interactions that focuses on evaluation and optimization of integrated/hybrid lethality and survivability technologies.

A reduction in the number of experiments needed to mature and integrate lethality and survivability concepts.

An enabling of high fidelity, three-dimensional simulations of complex projectile-target interactions to be conducted with fewer compromises in the model.

An acceleration of the discovery, refinement, development, and evaluation of innovative lethality and survivability technologies to attain the Army's technology readiness goals for FCS and beyond, i.e., Army 2010+.

A better understanding of interaction between various signatures and the development of innovative signature management techniques

Assist the technology in developing the ability to present different signatures in real time and change perception of adversaries' visualization of battle space

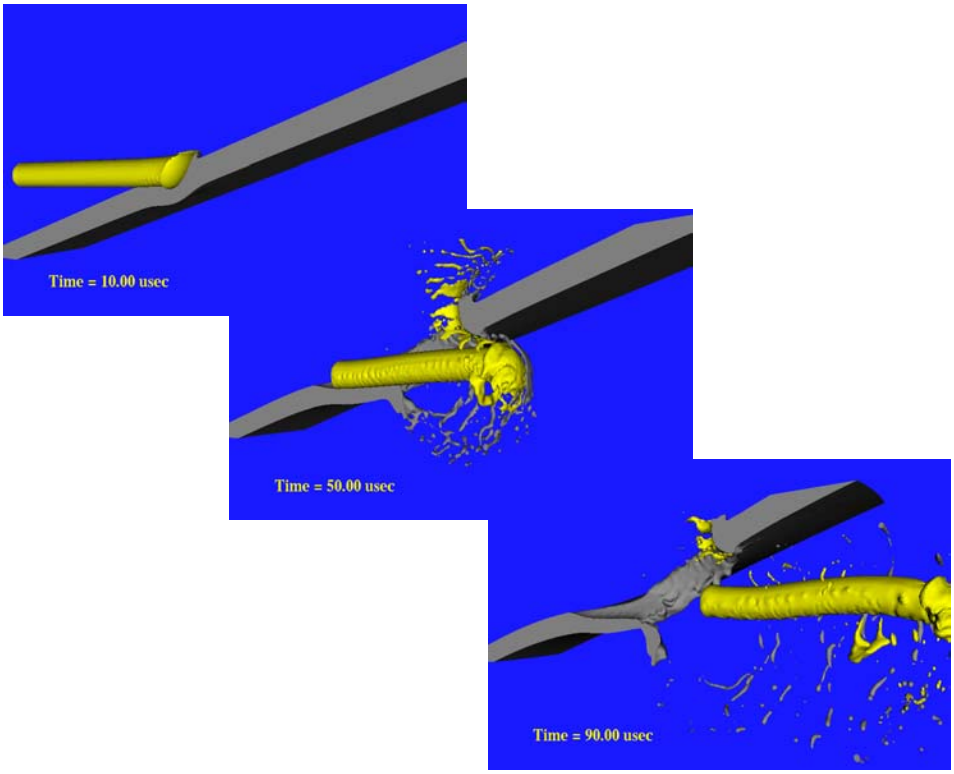
What could you accomplish if computers were 10x cheaper?

Lower cost HPC systems would permit deployment of large scalable systems (10x more processors). Furthermore, lower cost HPC systems would accelerate the deployment rate/number of processors and associated memory available to researchers in secure HPC environments. The majority of the HPC applications for this project address classified problems, so a significant increase in classified capability through the acquisition of lower cost processors/systems would help to address the computational requirements of this project. Since current shock physics software has demonstrated near liner scalability, researchers could readily exploit mega-processor HPC systems. Similarly, current electromagnetic effects software has also demonstrated near linear scalability. Computers with this capacity would allow researchers to model complex penetrator-target interactions and complex signature modeling approaches with unprecedented resolution and fide

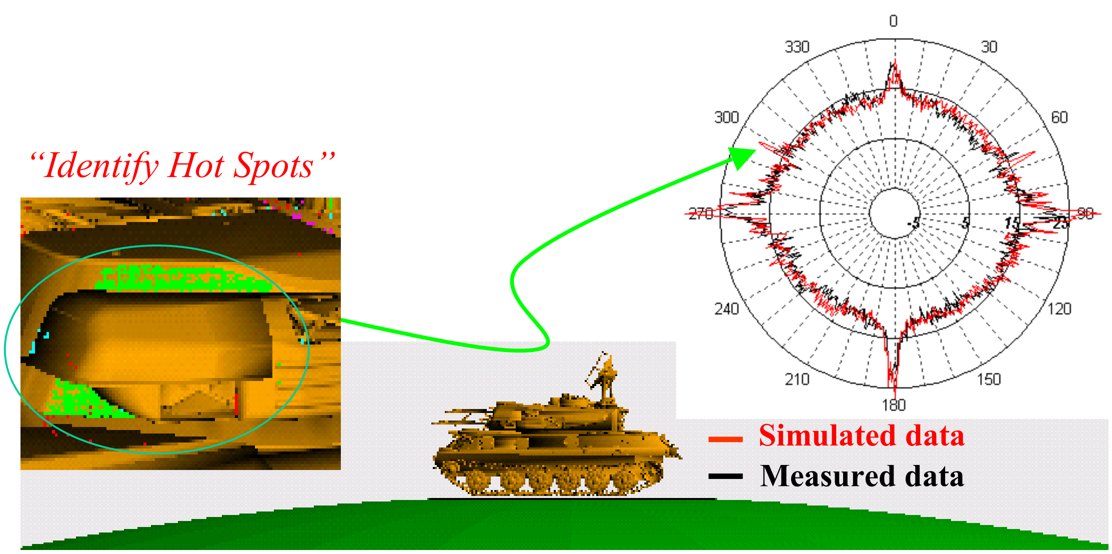
What could you accomplish if computers were 10x easier to program?

(All three of the above questions to be answered relative to progress greater than that expected from Moore's Law)

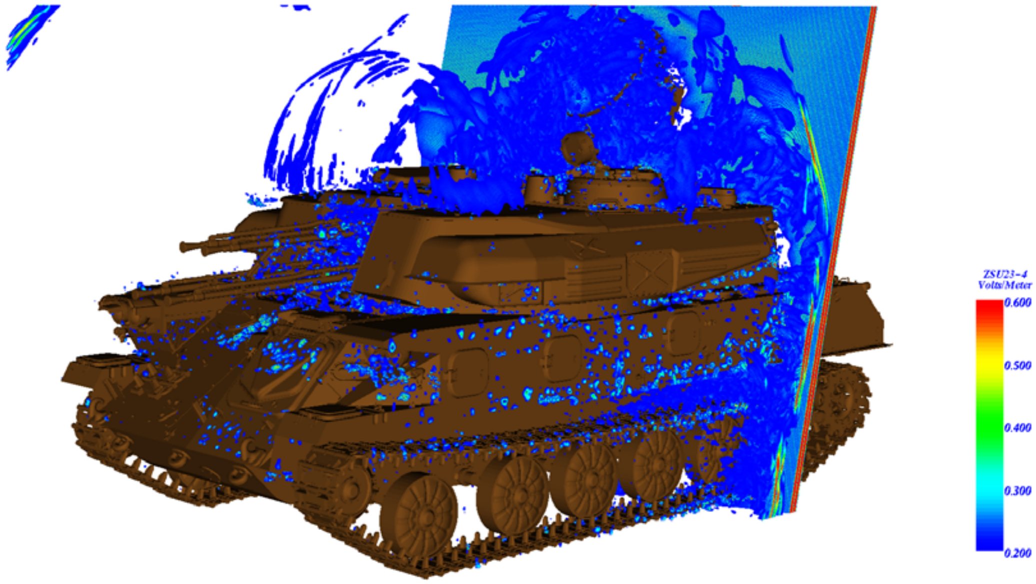
Today's application software for modeling solid dynamics problems involving large deformations, high strain rates, and shock wave propagation in multiple materials are well suited for exploiting scalable HPC systems. HPC systems that were easier to program would enable faster implementation of advanced material and failure models. This will enable researchers to address practical signature management issues involving multidisciplinary optimization approaches.



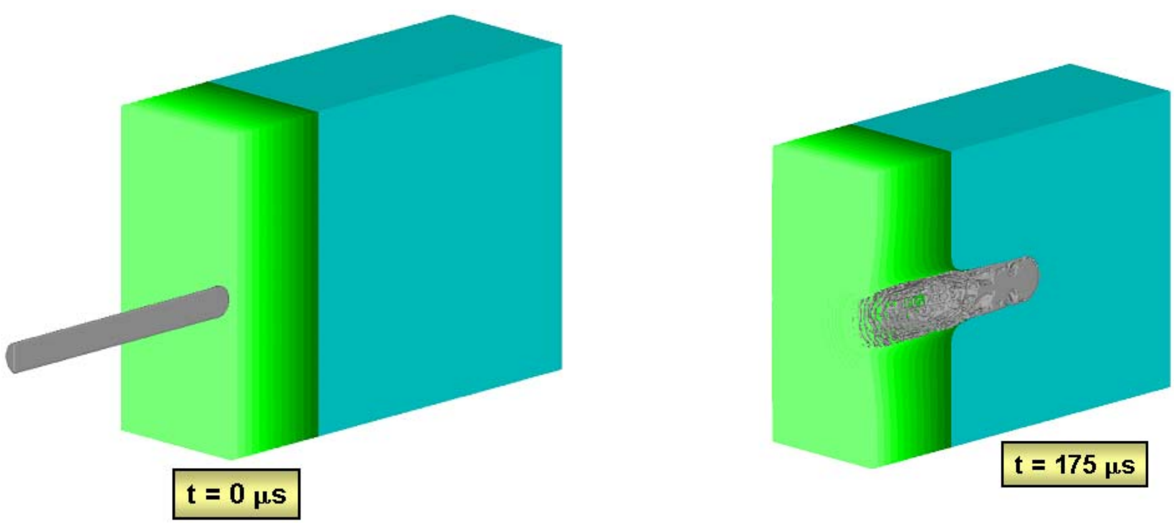
Modeling Complex Projectile-Target Interactions



Modeling Radar Cross Section in the Presence of the Ground Plane



ZSU combat vehicle subjected to an electromagnetic pulse at 10 GHz.



Modeling Functionally Graded Materials (FGM)

PROJECT NAME: **ARMY FUTURE COMBAT SYSTEMS**

HPC COMPUTATIONAL REQUIREMENTS

HPC System	Major Software Application	Sustained (Really Peak) Tera op/s Required for Each Fiscal Year					Security Classification
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012	
Scalable	ALEGRA	0.0684932	0.091324	0.091324	0.182648	0.365297	Unclassified
	CTH	0.0684932	0.091324	0.091324	0.182648	0.365297	Unclassified
	Paradyn	0.0684932	0.091324	0.091324	0.182648	0.365297	Unclassified
	Pronto	0.0684932	0.091324	0.091324	0.182648	0.365297	Unclassified
	Zapotec	0.0684932	0.091324	0.091324	0.182648	0.365297	Unclassified
Scalable	ALEGRA	5.0092006	7.871601	9.541334	19.08267	38.16534	Secret
	CTH	17.890002	28.11286	34.07619	68.15239	136.3048	Secret
	Paradyn	0.3578	0.562257	0.681524	1.363048	2.726096	Secret
	Pronto	0.3578	0.562257	0.681524	1.363048	2.726096	Secret
	Zapotec	0.3578	0.562257	0.681524	1.363048	2.726096	Secret
Scalable	ALEGRA	2.8538813	3.669276	4.484671	8.969341	17.93868	Dedicated
	CTH	4.2808219	5.503914	6.727006	13.45401	26.90802	Dedicated
	Paradyn	0.2853881	0.366928	0.448467	0.896934	1.793868	Dedicated
	Pronto	0.2853881	0.366928	0.448467	0.896934	1.793868	Dedicated
	Zapotec	0.2853881	0.366928	0.448467	0.896934	1.793868	Dedicated
Scalable	PFDTD	0.134722	0.233982	0.392602	0.785204	1.570408	Unclassified
	PFEM	0.134722	0.233982	0.392602	0.785204	1.570408	Unclassified

PROJECT NAME: **ARMY FUTURE COMBAT SYSTEMS**

HPC COMPUTATIONAL REQUIREMENTS

HPC System	Major Software Application	Sustained (Really Peak) Tera op/s Required for Each Fiscal Year					Security Classification
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012	
	PFISC	0.134722	0.233982	0.392602	0.785204	1.570408	Unclassified
	TEMPUS	0.134722	0.233982	0.392602	0.785204	1.570408	Unclassified
	XPATCH	0.134722	0.233982	0.392602	0.785204	1.570408	Unclassified
Scalable	PFDTD	0.965068	1.45936	1.997774	3.995548	7.991096	Secret
	PFEM	0.965068	1.45936	1.997774	3.995548	7.991096	Secret
	PFISC	0.965068	1.45936	1.997774	3.995548	7.991096	Secret
	TEMPUS	0.965068	1.45936	1.997774	3.995548	7.991096	Secret
	XPATCH	0.965068	1.45936	1.997774	3.995548	7.991096	Secret
Scalable	PFDTD	0.031203	0.069633	0.153728	0.614912	2.459648	Dedicated
	PFEM	0	0	0	0	0	Dedicated
	PFISC	0.031203	0.069633	0.153728	0.614912	2.459648	Dedicated
	TEMPUS	0	0	0	0	1.4891	Dedicated
	XPATCH	0.031203	0.069633	0.153728	0.614912	2.459648	Dedicated

PROJECT NAME: **ARMY FUTURE COMBAT SYSTEMS**

HPC MEMORY REQUIREMENTS

HPC System	Major Software Application	Typical Memory (GB total per run) Needed Each Fiscal Year					Security Classification
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012	
Scalable	ALEGRA	64	128	128	256	512	Unclassified
	CTH	32	64	64	128	256	Unclassified
	Paradyn	8	16	16	32	64	Unclassified
	Pronto	8	16	16	32	64	Unclassified
	Zapotec	40	80	80	160	320	Unclassified
Scalable	ALEGRA	512	768	1024	2048	4096	Secret
	CTH	256	384	512	1024	2048	Secret
	Paradyn	64	96	128	256	512	Secret
	Pronto	64	96	128	256	512	Secret
	Zapotec	320	480	640	1280	2560	Secret
Scalable	ALEGRA	512	768	1024	2048	4096	Dedicated
	CTH	256	384	512	1024	2048	Dedicated
	Paradyn	64	96	128	256	512	Dedicated
	Pronto	64	96	128	256	512	Dedicated
	Zapotec	320	480	640	1280	2560	Dedicated

Project Name: **Army Future Combat Systems**

HPC Software Requirements (Applications Overview)

Major Software Application	Computational Technology Area	Short Code Description	Code Size *(Lines of code)*	Rate of Change *(e.g., Fraction of code that changes each year)*	Languages Used *(Fraction of each)*
ALEGRA	CSM	Finite Element ALE	Several Million	25%	C++, C, Fortran
CTH	CSM	Finite Volume Euler	Several Million	5%	Fortran, C
Paradyn	CSM	Finite Element Lagrange	Several Million	5%	Fortran, C
Pronto	CSM	Finite Element Lagrange	Several Million	0%	Fortran
Zapotec	CSM	Coupled Euler+Lagrange	Several Million	25%	Fortran, C

			Key (for Computational Technology Area types)
			Computational Structural Mechanics (CSM)

PROJECT NAME: **ARMY FUTURE COMBAT SYSTEMS**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW (CONT.)

Major Software Application	Programming Models Used	Performance Critical Computation *(e.g., A linear solver or a sorting routine)*	Time to solution requirements *(S/W development time line requirements)*	Time to Solution Requirements *(Problem set-up time line requirements)*	Time to Solution Requirements *(Wall clock execution time for specified problem)*	Time to Solution Requirements *(Post run analysis time line requirements)*	Special Characteristics of Code *(e.g., Is it CPU performance, memory access, or I/O bound)*
ALEGRA	Message Passing	Explicit Time Integration	20-30 Work Yrs/Yr	2-4 Days	4-14 Days	1-2 Days	CPU & Memory Bound
CTH	Message Passing	Explicit Time Integration	2-3 Work Yrs/Yr	1-2 Days	4-14 Days	1-2 Days	CPU & Memory Bound
Paradyn	Message Passing	Explicit Time Integration	5-10 Work Yrs/Yr	1-2 Days	2-5 Days	1 Day	CPU & Memory Bound
Pronto	Message Passing	Explicit Time Integration	Fixed	1-2 Days	2-5 Days	1 Day	CPU & Memory Bound
Zapotec	Message Passing	Explicit Time Integration	5-10 Work Yrs/Yr	2-4 Days	4-14 Days	2-4 Days	CPU & Memory Bound

PROJECT NAME: **ARMY FUTURE COMBAT SYSTEMS**

HPC HARDWARE REQUIREMENTS (CURRENT SYSTEM BASELINE)

HPC System	Number of Computational Nodes Used by Typical Application	Number of Processors Per Node	Processor Characteristics *(e.g., Type, nominal speed)*	Memory Per Processor	I/O System Characteristics	Total Secondary Storage	Interconnect Characteristics *(Local and aggregate bandwidth, latency)*
Scalable U	64 Total Processors	All	RISC 10 GFLOPS	1 GB		32 GB/run	
Scalable S	512 Total Processors	All	RISC 10 GFLOPS	1 GB		256 GB/run	
Scalable D	256 Total Processors	All	RISC 10 GFLOPS	1 GB		256 GB/run	

PROJECT NAME: **ARMY FUTURE COMBAT SYSTEMS**

HPC SOFTWARE REQUIREMENTS (APPLICATION SCALABILITY)

Major Software Application	Scalability of Code for Fixed Problem Size	Scalability of Code for Increasing Problem Size *(Fixed Workload per Processor)*	Efficiency of Code on Target Platform
ALEGRA	N/A	Linear	85%
CTH	N/A	Linear	85%
Paradyn	N/A	Linear	Unknown
Pronto	N/A	Linear	85%
Zapotec	N/A	Unknown	Unknown

PROJECT NAME: **ARMY FUTURE COMBAT SYSTEMS**

HPC SOFTWARE REQUIREMENTS (DETAIL)

Major Software Application	Memory Access Pattern *(e.g., Stride 1 or random)*	Computation Profile *(e.g., Volume of integer v. floating point computation)*	I/O Requirements *(e.g., Volume and rate)*	Communication Characteristics *(e.g., Describe message size and rate)*	Inherent concurrency *(e.g., Number of independent threads in main body of calculation)*
ALEGRA					
CTH					
Paradyn					
Pronto					
Zapotec					

C-8 ELECTROMAGNETIC WEAPONS DEVELOPMENT

Project Description

Use of electromagnetic energy in a weapons device has been the dream of many war planners for decades. That dream is about to become a reality with the development of the airborne laser (ABL) for ballistic missile defense within the next few years. In addition, the Air Force is investigating the use of other electromagnetic devices for future use in disabling battlefield electronics and providing alternatives to lethal weaponry in denying enemy access to specified areas.

The ABL is a high-power, directed-energy weapon system that will be carried aboard a 747-400 freighter aircraft. Planned ABL system components include a detection subsystem for locating plume radiation from the launch site of an enemy ballistic missile, a low-power laser illumination subsystem for tracking the missile, and a high-power directed-energy chemical oxygen iodine laser system to destroy the missile in flight.

Three complementary facets of ABL design are being investigated in this project. First, laser power, laser gain, and laser gain medium quality are being predicted using simulations of the chemical laser system that generates the high-power laser. The second concerns the modeling of the strength, distribution, and spatial spectrum of stratospheric turbulence. The third effort then uses information about stratospheric turbulence to model optical propagation phenomena and the effects of these phenomena on optical tracking and adaptive optics systems for the laser.

The Air Force is modeling high-power microwave devices to improve their output characteristics for future battlefield use. Several different kinds of microwave devices are being studied.

What is the real life problem?

There are three major ABL design areas impacted by the HPC work described here:

1. Enhancement of the performance of chemical laser devices to enable their use in a variety of DoD missions, including ABL,
2. ABL operational adaptive-optics algorithm specification and design, and
3. ABL atmospheric operational decision aid.

The first of these requires detailed gas dynamic and electromagnetic modeling of the laser cavity and its output given the input to that cavity, which consists of species and concentrations produced by the chemical laser. The second involves three successive steps that must be carried out for optimal system operation: 1) characterize atmospheric optical properties; 2) model laser propagation through the atmosphere; and 3) develop adaptive optics compensation algorithms for ABL design based on laser propagation modeling. These steps are required to simulate the effects of atmospheric turbulence on the ability to deliver lethal fluency to a target many kilometers from the ABL and to use a combination of carefully chosen experiments and computer simulations to determine how well the degrading effects of the atmosphere can be alleviated by appropriate adaptive optical compensation. The third item, the operational decision aid, involves developing a methodology for predicting atmospheric turbulence during ABL deployment so that optimal missile-engagement strategies may by implemented. This component will be integrated with an operational weather forecast model to predict locations of intense atmospheric optical turbulence.

In order to achieve high power in a compact radio-frequency (RF) microwave device, it has to be operated at high current and high space charge. Consequently, these devices are highly non-linear. In addition, these devices have complex geometries with small features that require detailed modeling in order to produce accurate results.

What is the National Security Impact of (solving and not solving) the problem?

The ABL represents the boost-phase component of a layered U.S. ballistic missile defense system. Chemical lasers provide the ability to impact a target at the speed of light and at large ranges; this provides the ability to hit and kill intercontinental and theater range ballistic missiles, satellites, supersonic cruise missiles, and a variety of other targets. ABL is designed to destroy enemy missiles in the first 3 to 5 minutes after launch, and in so doing, address multiple ballistic missile systems with different operational ranges. By destroying a missile in its boost phase, the ABL will scatter debris from a successfully engaged missile over enemy territory so that it does not threaten the missile's intended target area.

ABL system design and performance evaluation is critically dependent on the capability to compute the performance of the laser and the effects of turbulence on optical propagation of the laser and adaptive optics used to position it. Reduced ABL effectiveness and/or non-optimal deployment will result if we do not properly characterize the impact of stratified atmospheric turbulence on optical propagation. In addition, once the impact of turbulence at operational altitudes is understood, the ability to forecast turbulence strength will be an important operational asset.

RF microwave weapons will be designed to work against hostile battlefield electronics.

Given the importance of electronics on the modern battlefield, the ability to target these electronics will give the U.S. a distinct advantage against technically advanced adversaries. The effect of RF radiation on the electronics can vary from transient disruption to destruction of the hardware, depending on the power and frequency of the radiation. Furthermore, since only electronic equipment is affected, RF radiation can be applied in a non-lethal manner, giving the war fighter a broader range of options for dealing with opponents of varying sophistication and strength. The generation of electromagnetic radiation is critical to the DoD's advanced RF weapon effort.

Why is HPC needed to address the problem?

Resolving spatial and temporal structure of chemically reacting gas flows with coupled complex physics found in chemical lasers incurs a large computational cost. The physical phenomena are non-linear and the device geometries are very intricate, both resulting in very large computational requirements. Simulations of chemical lasers, therefore, tend to require the types of computing capability afforded by HPC.

Understanding the nature of turbulence in the lower stratosphere and upper troposphere; deducing its effect on optical propagation and ultimately ABL performance; and developing techniques for forecasting likely locations and strength of optically significant atmospheric turbulence embodies three cutting-edge research fields—all of which require the significant computational resources provided by HPC assets. The reasons these components require beyond-state-of-the-art knowledge can be understood when one considers the nature of turbulence in this part of the atmosphere and current methods for quantifying it.

Stable stratification, or density layering with heavy air positioned beneath lighter air, results in the organization of turbulent patches when they are created by wind shear or overturning gravity waves. These turbulent patches can expand freely in horizontal layers, but are constrained vertically to only 100 meters to a few kilometers in depth.

Balloon measurements represent the highest-vertical-resolution technique for sampling these layers, however because they only offer a one-dimensional (1-D) glimpse of the atmosphere along the balloon's trajectory, detailed knowledge of the evolution, 3-D dynamics, and turbulence statistics of the layers is not possible. Radar measurements can provide some 3-D turbulence data continuously in time, but the resolution of this data is not sufficient to adequately characterize stratospheric turbulence at all length scales. Simulation techniques are necessary to develop a fully 3-D and temporal understanding of the individual turbulence layers that occur in the atmosphere, but the degree of turbulence attainable with modern supercomputers is far below those occurring in the atmosphere. Therefore, in order to learn from the simulation results, we must exploit the scaling behavior exhibited by the solutions so that we may extrapolate to atmospheric turbulence levels. We must also use atmospheric measurement results to validate and provide context for the numerical solutions.

Despite progress, the challenge of atmospheric turbulence simulation of the upper troposphere and the lower stratosphere is staggering. Current forecast models do not even resolve the individual turbulence layers, which means either extremely efficient methods must be developed which faithfully describe every aspect of stratified mixing layers, or significant advances in computational hardware must occur. Significant increases in computer hardware and/or advances in algorithm performance are needed just to simulate combined dynamics of fundamental processes, such as multiple interacting gravity waves, gravity waves in shear, or coupled shear layers.

For microwave devices, simulation of sources from first principles (Maxwell's equations and relativistic Lorenz's force law) requires electrodynamics and charged particle dynamics with second-order time and space accuracy. These simulations accurately predict RF-production and antenna gain. Such simulations require major computational resources. Thus, the ICEPIC code is able to effectively use hundreds of nodes on the massively parallel computers at the DoD high performance computing centers.

What do you need to accomplish with HPC between 2002 and 2012?

For the ABL as well as advanced laser devices, the performance of existing chemical laser technology must be improved by increasing laser power, volumetric and chemical efficiencies, and decreasing system weights. HPC will also be used to enable the development of alternate gas laser technologies, most probably electrically driven. The development of electrically driven gas laser technology directly interfaces with current efforts to drastically improve the power and efficiency of onboard power generation systems in DoD weapons platforms. For ABL specifically, improved statistical and phenomenological models of turbulence in stratified shear flows must be developed with the intent of upgrading the way that turbulence is represented in the simulation of optical propagation. Fundamental and practical limits on the ability of adaptive optics to compensate for turbulence effects on directed energy laser weapons must be developed, algorithmic and hardware techniques for improving laser weapons system performance must be explored, and a capability to forecast optically significant stratospheric turbulence using mesoscale weather code outputs must be developed. To better simulate future microwave weaponry devices, advanced parallel plasma physics software for compact HPM, W-band source design for airborne ADT, plasma processing, RF breakdown, hypersonic propulsion and drag reduction, charged particle beams, space applications, and beamed RF power must be developed.

What do you need (H/W & S/W) from the HPC Community between 2002 and 2012 to solve your problem?

All components of this project require faster and more processors. In addition, faster message passing both internal to and between system nodes is very important. For some types of calculations, larger memories per processor (both cache and main memory) are required. In addition, hardware systems are required for which higher percentages of peak performance are realized. In the software arena, the development of multi-resolution spectral techniques is required so that the cost advantages of adaptive-mesh algorithms may be combined with the resolving ability of spectral methods. Also needed is development of a parallel algorithm for the tar command so that massive parallel data sets may be rapidly archived without consuming large numbers of system addresses. Increased performance of all applications software through hardware improvement or software (compiler) improvement is necessary. New algorithms are also needed that add higher fidelity physics and/or hybrid models that allow the simulation of higher density plasmas.

Project Vision

What is the mission vision?

In high-power laser development the vision is to improve the performance of current chemical laser technology and enable advances in gas laser technology in general to facilitate the use of high power lasers as weapon systems. In modeling the ABL system, the specific vision is to have a capability of simulating at least 1 second of ABL system operation per day accounting for the effects of illuminator partial coherence. This requires the ability to provide accurate statistical and phenomenological models of stratospheric turbulence that can be used in these laser weapons system simulations.

To accurately forecast optically significant atmospheric turbulence, microscale parameterizations must be determined from the output of mesoscale weather codes. The goal is to obtain an improved turbulence forecasting capability that will be included as a part of a laser weapons system atmospheric decision aid.

The virtual prototyping concept for microwave weapons devices substantially shortens the design cycle of microwave sources that are capable of providing useful militarily effects.

The vision is to be able to simulate one nanosecond in one second of real time. This would allow the design of high-power microwave sources in near-real time. Currently, it takes one hour of real time to simulate one nanosecond of real time.

What could you accomplish if computers were 10x more powerful?

Substantially greater physical detail would be added to the 3-D chemical laser simulations currently conducted. This would also facilitate the simulation of full-scale hardware, an important aspect of optimizing the performance of chemical lasers as systems. Increasing simulation times and numbers of statistically independent cases will produce better statistics on laser weapons system performance. More accurately modeling relevant physics using a finer/larger computational mesh will result in predictions of greater accuracy. Because the cost of a turbulence simulation is proportional to the cube of the Reynolds number, an important flow parameter, an order of magnitude greater computing power is required simply to double that parameter and make the calculation roughly twice as accurate. Though this brute force technique of obtaining greater accuracy is of value, it is clear that spectral-method algorithm improvements are required for significant increases in accuracy. Such algorithms would have a profound impact on computational fluid dynamics in general, not just turbulence modeling. Higher resolution weather and optical turbulence forecasts could be carried out, but evaluation and possibly modification of current mesoscale-model sub-grid-scale parameterizations would be required.

Simulations of high-power microwave devices have grown in fidelity with the addition of resolution and/or more detailed physics models to fill available computational resources. This is true for two reasons: first, microwave tubes have far more non-linear complexities than was first realized, and, second, electromagnetic particle-in-cell codes formerly gave only qualitative explanations on how microwave tubes work, but now they are producing quantitative results that match experiment. An additional factor of ten speed-up in computers would allow delivery of more accurate results in a timely manner.

What could you accomplish if computers were 10x cheaper?

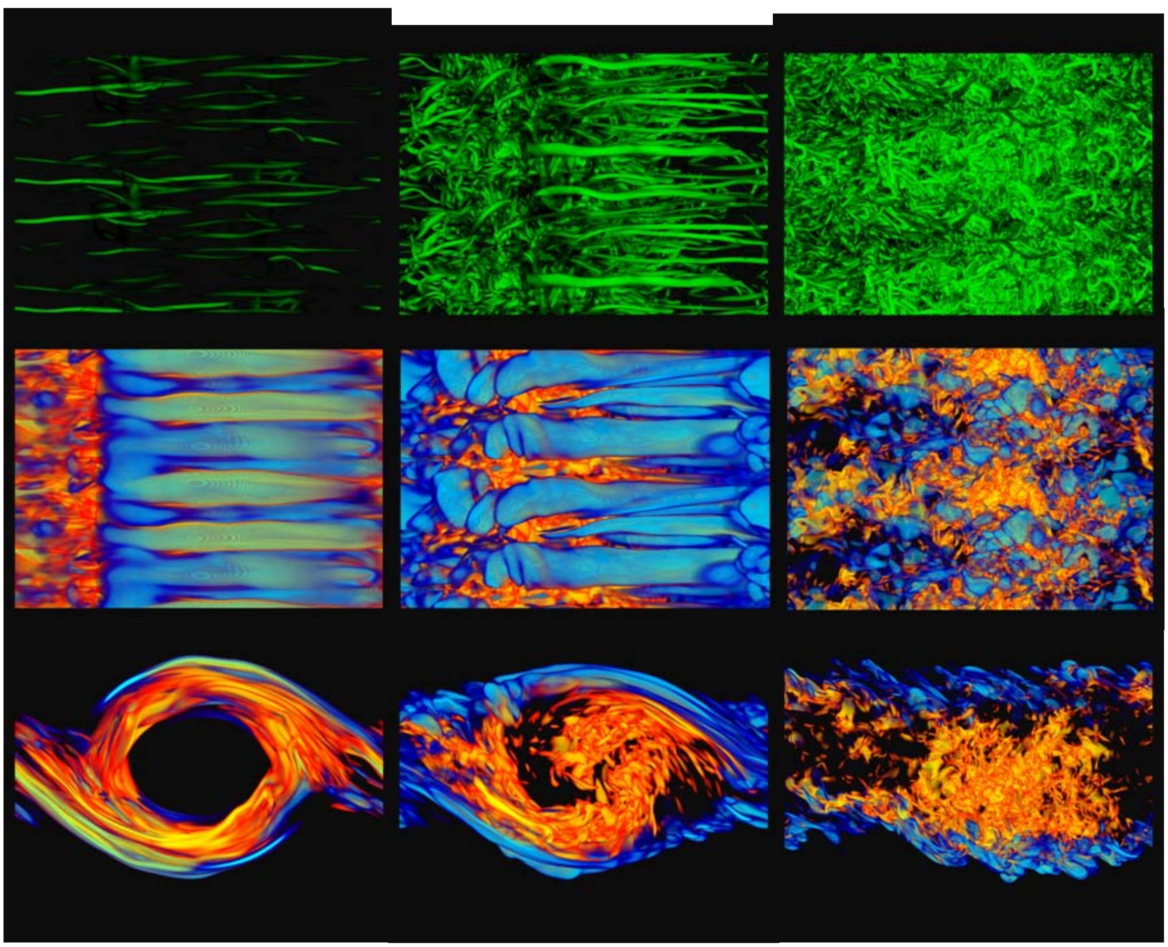
Assuming networking design kept pace, a factor of 10 savings in computer cost translates directly to a factor of 10 increase in computer power. One possibility of 10x cheaper hardware is the acquisition of in-house computing as opposed to using DoD or other centralized systems. For this to be realistic, however, maintenance and archival storage systems would also have to exhibit the same price reduction, and even then it would be more feasible to house such systems in house if the cost were 100x rather than 10x cheaper. Unless this happened, it would still be a far better alternative to use centralized systems, particularly if the factor of 10 savings was used to acquire 10x as much computational capability. In this case, the answers to the previous question apply.

What could you accomplish if computers were 10x easier to program?

Assuming performance does not take a back seat to coding simplicity, easier coding would be beneficial. However, it has not been a general experience that attempts to streamline coding through automatic serial code parallelization, for example, have been successful, largely because the tools that have been produced do not generate optimized code. A good programmer often seems to be able to write code that runs twice as fast as the laborsaving tools. On the other hand, creation of a more user-friendly programming environment would greatly facilitate the improvement of physical fidelity within existing models and allow experimentation with new algorithms much more easily.

The images below depict numerical simulations of the turbulent dynamics within a single cloud billow in the stratosphere. These turbulence simulations represent the highest resolution stratified turbulence simulations ever conducted. Images are viewed from above and from the side for three different times as the flow evolves.

The green images show the turbulence morphology by depicting vortex tubes in the flow. The blue and orange images show the square of the index-of-refraction gradient in blue and the square of the velocity gradient in orange. The images clearly demonstrate the partitioning of the flow into regions where optical turbulence effects are important (blue regions) and where mechanical turbulence dominates (orange regions). The bottom row of images shows the view from the side; this row is most easily identifiable as a single billow from the cloud images. Visualizations like these show that intense mixing by mechanical turbulence reduces optical effects, but in the regions between turbulent and quiet motion of the atmosphere intense optical turbulence persists for long periods of time.



PROJECT NAME: **ELECTROMAGNETIC WEAPONS DEVELOPMENT**

HPC COMPUTATIONAL REQUIREMENTS

HPC System	Major Software Application	Number of CPU Hours (or sustained Tera op/s) Required for Each Fiscal Year					Security Classification
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012	
A	A-1	XX	XX	XX	XX	XX	XX
	A-2	XX	XX	XX	XX	XX	XX
IBM P3	optical propagation	395,000	750,000	1,500,000	2,500,000	5,000,000	Unclassified
SGI Origin 3K	Custom CFD	600,000	1,200,000	2,400,000	3,600,000	5,000,000	Unclassified
Compaq SC40/SC45	ICEPIC	300,000	750,000	1,500,000	2,000,000	2,500,000	Unclassified
IBM Netfinity	ICEPIC	100,000	300,000	1,000,000	2,000,000	2,500,000	Unclassified
IBM SP/P3	ICEPIC	500,000	1,000,000	2,000,000	2,500,000	3,000,000	Unclassified
ASC IBM SP3	GASP	400,000	400,000	600,000	900,000	1,350,000	Unclassified
MHPCC IBM SP3	GASP	150,000	150,000	225,000	337,500	506,250	Unclassified
ASC Compaq SC45	GASP	150,000	150,000	225,000	337,500	506,250	Unclassified
ERDC Compaq SC45	GASP	150,000	150,000	225,000	337,500	506,250	Unclassified
Compaq ES45	triple_zyx.f	200,000	600,000	2,000,000	3,700,000	5,500,000	Unclassified
Cray T3E	triple_zyx.f	750,000	850,000	0	0	0	Unclassified
IBM SP/P3	les_mpi.f	300,000	400,000	600,000	1,200,000	2,000,000	Unclassified

PROJECT NAME: **ELECTROMAGNETIC WEAPONS DEVELOPMENT**

HPC MEMORY REQUIREMENTS

HPC System	Major Software Application	Typical Memory (GB total per run) Needed Each Fiscal Year				
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012
A	A-1	XX	XX	XX	XX	XX
	A-2	XX	XX	XX	XX	XX
IBM P3	optical propagation	16	32	64	128	128
SGI Origin 3K	Custom CFD	see 017 req survey				
Compaq SC40/SC45	ICEPIC	48	64	128	192	256
IBM Netfinity	ICEPIC	24	32	64	128	192
IBM SP/P3	ICEPIC	24	64	128	192	256
ASC IBM SP3	GASP	64	128	256	512	1024
MHPCC IBM SP3	GASP	64	128	256	512	1024
ASC Compaq SC45	GASP	64	128	256	512	1024
ERDC Compaq SC45	GASP	64	128	256	512	1024
Compaq ES45	triple_zyx.f	50-200	70-280	100-400	190-760	280-1000
Cray T3E	triple_zyx.f	50-200	70-280	100-400	190-760	280-1000
IBM SP/P3	les_mpi.f	50-100	70-140	80-200	120-250	200-400

PROJECT NAME: **ELECTROMAGNETIC WEAPONS DEVELOPMENT**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW)

Major Software Application	Computational Technology Area	Short Code Description	Code Size *(Lines of code)*	Rate of Change *(e.g., Fraction of code that changes each year)*	Languages Used *(Fraction of each)*
A-1	CFD	xx	xx	xx	xx
A-2	CSM	xx	xx	xx	xx
Optical propagation	CEA	optical prop	16,000	5 to 10%	Fortran
Custom CFD	CFD	Turbulence Sim	?	?	Fortran
ICEPIC	CEA	EM PIC	45000	5%	C
ICEPIC	CEA	EM PIC	45000	5%	C
ICEPIC	CEA	EM PIC	45000	5%	C
GASP	CFD	Computational fluid dynamics coupled to chemistry and optics.	350,000	10-15%	31% Fortran 77, 69% C++
triple_zyx.f and support routines	CFD	Spectral Boussinesq Solver	8,000 + 20,000	10%	Fortran (100%)
les_mpi.f	CFD	Spectral LES Solver	10,000	10-20%	Fortran (100%)
triple_zyx.f support scripts	CFD	backup /housekeeping	10,000	25%	Perl (70%), csh/sh (30%)
			Key (for Computational Technology Area types)		
			Computational Structural Mechanics (CSM)		

PROJECT NAME: **ELECTROMAGNETIC WEAPONS DEVELOPMENT**					
HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW)					
Major Software Application	**Computational Technology Area**	**Short Code Description**	**Code Size** *(Lines of code)*	**Rate of Change** *(e.g., Fraction of code that changes each year)*	**Languages Used** *(Fraction of each)*
			Computational Fluid Dynamics (CFD)		
			Computational Electromagnetics and Acoustics (CEA)		

PROJECT NAME: **ELECTROMAGNETIC WEAPONS DEVELOPMENT**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW (CONT.))

Major Software Application	**Programming Models Used**	**Performance Critical Computation** *(e.g., A linear solver or a sorting routine)*	**Time to Solution Requirements** *(S/W development time line requirements)*	**Time to Solution Requirements** *(Problem set-up time line requirements)*	**Time to Solution Requirements** *(Wall clock execution time for specified problem)*	**Time to Solution Requirements** *(Post run analysis time line requirements)*	**Special Characteristics of Code** *(e.g., Is it CPU performance, memory access, or I/O bound)*
A-1	xx	xx	xx	xx	xx	xx	xx
A-2	xx	xx	xx	xx		xx	xx
optical propagation	SPMD	FFT	Varies	2 to 5 min	10min to 48hrs	30min to a few days	CPU and message passing
Custom CFD	?	FFT	?	?	?	?	?
ICEPIC	OO/procedural	linear solve	days to months	days to weeks	hours to days	hours to days	primarily CPU perf
ICEPIC	OO/procedural	linear solve	days to months	days to weeks	hours to days	hours to days	primarily CPU perf
ICEPIC	OO/procedural	linear solve	days to months	days to weeks	hours to days	hours to days	primarily CPU perf

PROJECT NAME: **ELECTROMAGNETIC WEAPONS DEVELOPMENT**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW (CONT.))

Major Software Application	Programming Models Used	Performance Critical Computation *(e.g., A linear solver or a sorting routine)*	Time to Solution Requirements *(S/W development time line requirements)*	Time to Solution Requirements *(Problem set-up time line requirements)*	Time to Solution Requirements *(Wall clock execution time for specified problem)*	Time to Solution Requirements *(Post run analysis time line requirements)*	Special Characteristics of Code *(e.g., Is it CPU performance, memory access, or I/O bound)*
GASP	xx	implicit time integration, i.e. jacobian generation and matrix inversion	8 hrs	320 hrs	156.25	80 hrs	memory access bound, slightly more so than CPU performance bound
GASP	xx	xx	xx	xx	xx	xx	xx
GASP	xx	xx	xx	xx	xx	xx	Xx
triple_zyx.f	SPMD	FFT	xx	xx	10-20 hours	3-5 hours	large cache, efficient I/o and archival storage access
les_mpi.f	SPMD	FFT, Pressure solver	xx	xx	10-20 hours	3-5 hours	large cache, efficient I/o and archival storage access

PROJECT NAME: **ELECTROMAGNETIC WEAPONS DEVELOPMENT**

HPC HARDWARE REQUIREMENTS (CURRENT SYSTEM BASELINE)

HPC System	Number of Computational Nodes Used by Typical Application	Number of Processors Per Node	Processor Characteristics *(e.g., Type, nominal speed)*	Memory Per Processor	I/O System Characteristics	Total Secondary Storage	Interconnect Characteristics *(Local and aggregate bandwidth, latency)*
A	xx	xx	xx	xx	xx	xx	xx
B	xx	xx	xx	xx	xx	xx	xx
IBM, P3	1 to 8	16	400 MHZ Power 3	512MB	?, Not an issue	?, Not an issue	Message passing is an issue
Compaq SC40/SC45	8	4	Alpha, 833 MHz	1GB	NA	16GB	xx
IBM Netfinity	16	2	Pentium III, 933 MHz	0.5GB	NA	16GB	xx
IBM SP/P3	2	16	Power 3, 375 MHz	0.5GB	NA	16GB	xx
IBM SP3	32 to 64	4	375 Mhz IBM Power3	1 Gb	xx	xx	150 mb/sec
Compaq ES45	32 to 64	4	1Ghz Alpha EV68	1 Gb	xx	xx	200 mb/sec

PROJECT NAME: **ELECTROMAGNETIC WEAPONS DEVELOPMENT**

HPC HARDWARE REQUIREMENTS (CURRENT SYSTEM BASELINE)

HPC System	Number of Computational Nodes Used by Typical Application	Number of Processors Per Node	Processor Characteristics *(e.g., Type, nominal speed)*	Memory Per Processor	I/O System Characteristics	Total Secondary Storage	Interconnect Characteristics *(Local and aggregate bandwidth, latency)*
Compaq ES45	50-125	4	xx	xx	O(Tbyte) per run	O(Tbyte) archival storage, O(100 Gbytes) system disks	local: 20-40 Gbytes/few minutes; internet: 5Mbits/sec
Cray T3E	201-1001	1	xx	xx	O(Tbyte) per run	O(Tbyte) archival storage, O(100 Gbytes) system disks	local: 20-40 Gbytes/few minutes; internet: 5Mbits/sec
IBM SP/P3	50-250	4	xx	xx	O(0.5 Tbyte) per run	O(0.5 Tbyte) archival storage, O(50 Gbytes) system disks	local: 20-40 Gbytes/few minutes; internet: 5Mbits/sec

PROJECT NAME: **ELECTROMAGNETIC WEAPONS DEVELOPMENT**

HPC SOFTWARE REQUIREMENTS (APPLICATION SCALABILITY)

Major Software Application	Scalability of Code for Fixed Problem Size	Scalability of Code for Increasing Problem Size *(Fixed Workload per Processor)*	Efficiency of Code on Target Platform
A-1	xx	xx	xx
A-2	xx	xx	xx
Optical Propagation	not good for small problems	approximately 100%	95 to 100 %
Custom CFD	?	?	?
ICEPIC	90%	95%	xx
ICEPIC	85%	88%	xx
ICEPIC	95%	95%	xx
GASP	220 for 256 processors	220 for 256 processors	86%
triple_zyx.f	xx	linear beyond 600 PE	25 -30%
les_mpi.f	xx	linear beyond 300 PE	new code, testing required

PROJECT NAME: **ELECTROMAGNETIC WEAPONS DEVELOPMENT**

HPC SOFTWARE REQUIREMENTS (DETAIL)

Major Software Application	**Memory Access Pattern** *(e.g., Stride 1 or random)*	**Computation Profile** *(e.g., Volume of integer v. floating point computation)*	**I/O Requirements** *(e.g., Volume and rate)*	**Communication Characteristics** *(e.g., Describe message size and rate)*	**Inherent Concurrency** *(e.g., Number of independent threads in main body of calculation)*
A-1	xx	xx	xx	xx	xx
A-2	xx	xx	xx	xx	xx
optical propagation	stride 1	mostly float	modest	1MB or larger	varies from one to several
Custom CFD	?	mostly float	very large	large	?
ICEPIC	stride	floating	xx	xx	xx
ICEPIC	stride	floating	xx	xx	xx
ICEPIC	stride	floating	xx	xx	Xx
GASP	stride 1	99%	200 mb/sec	Ave Size: 1 mb; Total bandwidth required: greater than 50 mb/sec	same as number of processors
GASP	xx	xx	xx	xx	xx
GASP	xx	xx	xx	xx	xx

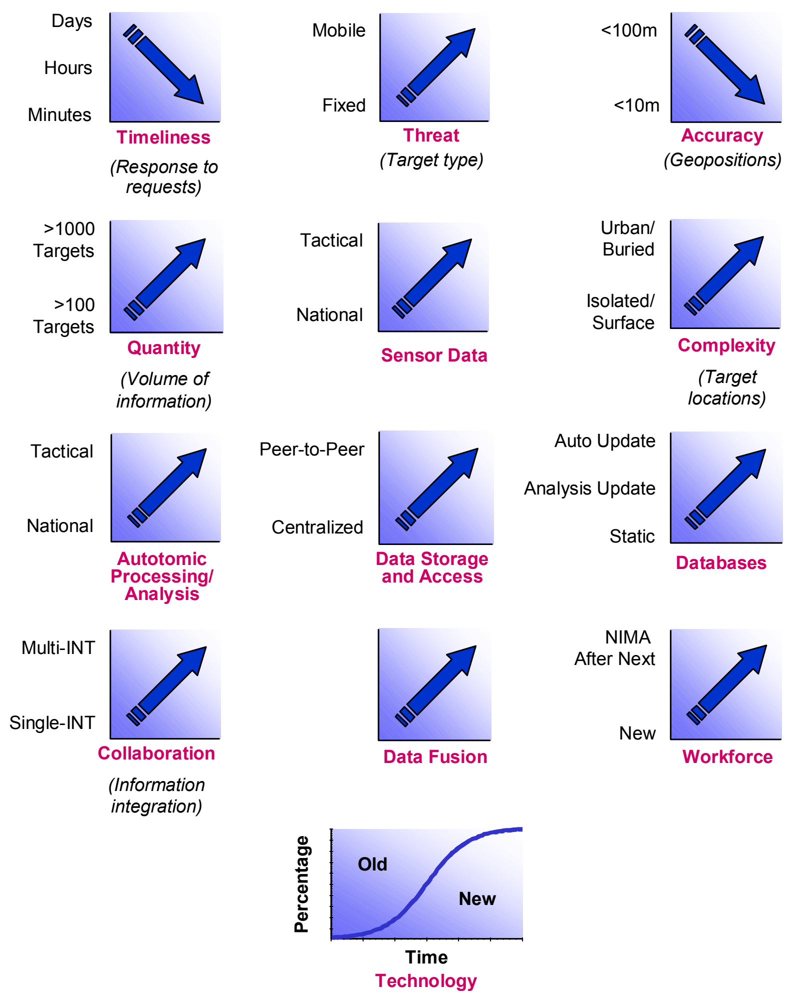
PROJECT NAME: **ELECTROMAGNETIC WEAPONS DEVELOPMENT**

HPC SOFTWARE REQUIREMENTS (DETAIL)

Major Software Application	Memory Access Pattern *(e.g., Stride 1 or random)*	Computation Profile *(e.g., Volume of integer v. floating point computation)*	I/O Requirements *(e.g., Volume and rate)*	Communication Characteristics *(e.g., Describe message size and rate)*	Inherent Concurrency *(e.g., Number of independent threads in main body of calculation)*
triple_zyx.f	stride 1	< 1/1000	50 Gigabyte each hour	each processsor transfers 10 Mwords to each of the other processors per hour (500 PE job)	same as number of processors used
les_mpi.f	stride 1	< 1/1000	50 Gigabyte each hour	each processsor transfers 10 Mwords to each of the other processors per hour (500 PE job)	same as number of processors used

C-9 GEOSPATIAL INTELLIGENCE

Data fusion and precision intelligence needs involving motion video, lidar, thermal, seismic, gravimetrical, and other imagery-derived phenomenologies are driving the need to solve technological challenges for geospatial intelligence derivation and delivery. Some of these are illustrated below.



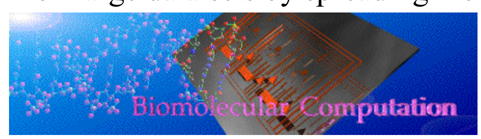
To generalize, two major facets arise that require high end computing investigation and solutions: 1) increase need to perform basic data analysis on various forms of geospatial data, 2) increased size and mix of potential data sources to be used for geospatial intelligence analysis.

Problem area (1) is a recognition that current data sets as they arrive independently, need some level of processing and pre-processing to allow for human involved analysis. These data sets can and do range from standard EO to IR, RADAR, multispectral and hyperspectral modalities. Pre-9/11 approximately 20% of these data sets can be processed for actual use. Post

9/11 the percentage rises with the inclusion of homeland security requirements and motion video availability. As new and novel data sets arrive for analysis (e.g. streaming video, commercial imagery, etc.), the types and amount of analysis will radically alter the amount of data that can be processed. Estimates before 9/11 for future and other NTM sensor suites show that NIMA's processing capability will drop to ~3% or less of all incoming data. After 9/11 with the advent of increased use of unmanned collection systems and new requirements to support homeland security applications, the processing capability will drop ***even lower***. More and more nformation will go un-analyzed due to lack of processing capability. With geospatial intelligence being the reference intelligence for all national security missions, this will pose a shortfall in overall defense capability.

Problem area (2) involves the concept of multi-int analysis applied to heterogeneous data sets. These data sets will have varying accuracies applied to their data types and will need to be analyzed in an environment that will allow for novel data types to be extracted. This problem is typically identified as conflation, which is the involvement not only of heterogeneous data sets but of needing specialized skills for extractinguseful data from multiple overlapping data types.

High performance computing (HPC) through either a distributed, cluster or grid, optical or DNA (the characteristics of bio and DNA computers are well suited to image and geospatial analysis) computing concept is one that would augment work on large data sets by spreading the work suite over multiple processors. Such an idea requires data sets that are amenable to such processing, programs able to handle disparate computing capability (load balancing), management of large data sets being distributed to multiple processing areas and a final re-integration and QC of finished work. Large geospatial data sets are inherently amenable to being broken down into smaller sub-functional areas as seen through previous capability in mosaic work and in GIS areas for extracting new data from disparate data sets. Although no programs have been designed for geospatial work in a distributed HPC environment, a high degree of ROI can be expected, as data flows into such an environment will be continuous. An area to expand for investigation is the area of DNA computing. The characteristics of DNA computing devices that would be of most interest to NIMA, its clients, and its suppliers appear to be those associated with huge memories. The volume of data maintained by NIMA is vast and growing rapidly. Thus, the availability of devices that could store such volumes easily and cheaply, and that could recall data quickly and accurately, would be of significant value to the agency.



Since much of the imagery information required by NIMA customers does not require detailed accuracy, the possibility of error inherent in most DNA computer processes would not appear to be a major problem for NIMA in many of its processes. However, since biological processes tend to be slow, DNA computers will probably be of little value to NIMA in its goal to get information to its clients in a timelier manner. The use of DNA computers in cryptographic processes may also be of value to NIMA, as well as to several of its customers.

Looking at HPC through distributed computing capability has many benefits for analysis work as multiple, heterogeneous data sets are amenable to being broken into smaller pieces for processing. Being able to have pre-processed working spaces for geospatial intelligence analysis will take a processing burden off of individual workstations and distribute such work across a network of devices for processing needs. Such a system is inherently scalable and has redundancy built into it. The loss of a single node or computer does not bring the entire distributed set of computers down and recognition of that loss will allow for automated programs to re-route data for processing to still existing nodes. In a work environment this will allow for multiple jobs to be tasked throughout NIMA for simultaneous processing. Combined with network-based storage the final end result of such processing will be available for analysis extraction work anywhere on the network to those users assigned to do such work.

Thus, distributed HPC has multiple benefits for geospatial intelligence applications. 1) Use in pre-processing of data sets for basic analysis work by composing a much larger data environment out of heterogeneous data types, 2) Pre-processing for work in doing mosaics, data set combining, digital elevation modeling, and generating up a multi-dimensional viewing space out of large data sets, 3) Continuous processing of incoming data so as to make it available for either human or computer based extraction methods, 4) By enabling distributed processing and combining it with a distributed storage environment, work can be done on such a data set from any node on that network capable of displaying the resulting work using the proper analysis tools.

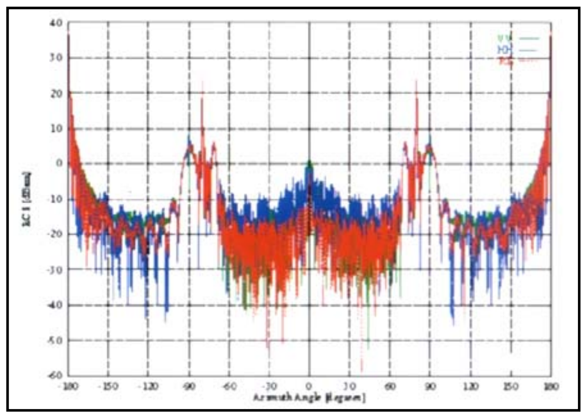
C-10 THREAT WEAPON SYSTEMS CHARACTERIZATION

Project Description

The DIA Threat Weapons System Characterization effort is an integral part of the Intelligence Community's mission of assessing threats to the United States, its allies, and its assets. Where high performance computing is concerned, these efforts currently focus on computational aerodynamics, signature prediction, and threat system performance modeling and simulation.

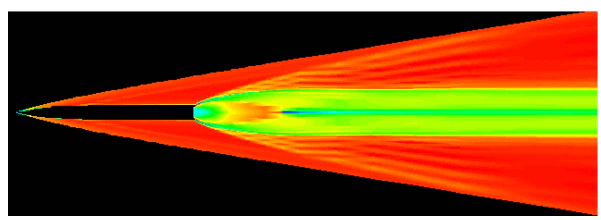
What is the real life problem?

U.S. operational forces, as well as many civilian assets that might be threatened by terrorists, depend upon accurate assessments of threat weapons so that appropriate defensive systems or countermeasure systems can be developed and fielded.



Developing responses to threat weapons whose technical sophistication and capabilities are on par with U.S. systems requires extremely detailed assessment and characterization of the threats. For example, the Missile Defense Agency requires not only accurate descriptions of the signatures of the missile threats its systems will engage, but also detailed estimates of the expected trajectories of those missiles.

What is the National Security Impact of (solving or not solving) the problem?



The U.S. cannot respond to threats to our national interests if we cannot characterize the threats. Systems being developed by the Missile Defense Agency, for example, depend upon accurate missile threat characterization (e.g., signatures, trajectories) to be successful. Other examples of intelligence issues in which solving the types of problems addressed by the HPCS affect national security include support to national warning systems (e.g., detection of ballistic missile attack), tactical missile launch detection systems, automatic target recognition systems, and battlefield surveillance systems.

Overall, the U.S. must respond to the increasingly technical battlefield of the future with high quality scientific characterizations of the scientific and technical threats on that battlefield.

Why is HPC needed to address the problem?

Without a doubt, the applications needed to address complex threat characterization issues require high performance computing. The primary reasons include:

- Since the Intelligence Community (IC) seldom has access to actual foreign hardware for wind tunnel testing or signature measurements, the only efficient and cost effective methods available to IC engineers are computational techniques. Uncertainties in the physical characteristics of threat weapons (and the subsequent need for variations of those definitions so statistical significance can be achieved) tend to generate prohibitive costs if estimates are based on hardware modeling and field-testing.
- Since aerodynamics characteristics and signature estimates must be developed for a complete set of desired values (i.e., many aspect angles, many different wavelengths, all likely angles of attack and Mach number, etc.), the timeliness and costs of field measurements are prohibitive.
- Some types of foreign hardware do not easily lend themselves to hardware testing. For example, very portable missiles that might be used by terrorists against commercial aircraft are so small they cannot be readily instrumented for wind tunnel testing over their actual flight regimes.
- There are so many foreign threat weapons that high performance computing offers the only possibility of evaluating each one. The costs and time that would be required to evaluate each significant threat using hardware testing methodologies are so large that they are absolutely unreasonable.
- As a Combat Support Agency, DIA must be capable of responding immediately with advanced scientific analyses of changing battlefield conditions. In other words, if a technical surprise occurs on a battlefield, DIA must be in a position to immediately apply all its scientific resources against that problem. High performance computing allows DIA to respond to battlefield crises.

What do you need to accomplish with HPC between 2002 and 2012?

Between 2002 and 2012, we expect the requirements for detailed assessments for threat weapons to grow tremendously. As the U.S. develops more systems that depend upon battlefield awareness to gain advantages over our adversaries, the quality and quantity of detailed threat characterizations will expand exponentially. We must improve not only the throughput of our high performance computing (to improve our ability to respond to changing threat conditions), but also the accuracy and scope of our threat characterization methodologies.

In the case of missiles, the extremely intensive methodologies required to estimate threat characteristics could only be executed on high performance computer systems.

What do you need (H/W & SW) from the HPC Community between 2002 and 2012 to solve your problem?

As at most organizations, MSIC's requirements for high performance computing enhancements are a balance between application software, system software, and system hardware. Besides obvious needs for improved application software (accomplished by research into improved methodologies, algorithms, and coding techniques), one of greatest needs is for improved batch scheduling systems.

What could you accomplish if computers were 10x more powerful?

In the case of DIA's HPCS, most of the applications are limited by the raw computing power available within each processor. We have not reached the point in our application developments at which performance figures are not limited by either processor speed or available memory. In fact for most cases, the quality of the methodologies used to address intelligence problems is limited by the computing power available to address those issues. As more powerful computers become available, not only will the speed at which the applications are executed increase, but the quality of the resulting intelligence product will also improve.

What could you accomplish if computers were 10x cheaper?

Since at DIA/MSIC the configuration of the HPCS is determined by maximizing the ratio of computing power to system costs, up to a point reducing the costs of computers would result in an equivalent increase in computing power. MSIC, however, is beginning to experience problems with the infrastructure associated with the HPCS (i.e., air conditioning and physical space). Decreasing the costs of computers, therefore, does not automatically equate to a capability to expand the size of the HPCS, since it may not be cost effective to increase the air conditioning and physical space. In an overall sense, MSIC would hope that the capability density (i.e., the ratio of speed to volume) would increase as the computer costs decrease.

What could you accomplish if computers were 10x easier to program?

In the case of applications on the HPCS, the issues associated with the software center on the basic analytical methodologies, the software algorithms associated with those methods, and the computer systems software. Clearly, as improvements are made in these areas, the quality of our intelligence support to critical customers (such as the Missile Defense Agency) will improve.

PROJECT NAME: **THREAT WEAPON SYSTEMS CHARACTERIZATION**

HPC COMPUTATIONAL REQUIREMENTS

HPC System	Major Software Application	Number of CPU Hours (or sustained Tera op/s) Required for Each Fiscal Year					Security Classification
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012	
Linux	Performance Modeling and Simulation	.03 T	.04 T	.07 T	.15 T	.2 T	Classified (SCI)
Linux	Signature Prediction	.04 T	.07 T	1 T	2 T	4 T	Classified (SCI)
Linux	Computational Aerodynamics	1.3 T	35 T	70 T	140 T	210 T	Classified (SCI)
SGI O2K	Performance Modeling and Simulation	.06 T	.03 T				Classified (SCI)
Sun E420	Signature Prediction	.1 T					Classified (SCI)
Sun E420	Computational Aerodynamics	.1 T					

PROJECT NAME: **THREAT WEAPON SYSTEMS CHARACTERIZATION**

HPC MEMORY REQUIREMENTS

HPC System	Major Software Application	Typical Memory (GB total per run) Needed Each Fiscal Year				
		FY 2003	FY 2005	FY 2007	FY 2010	FY 2012
ANY	Performance Modeling and Simulation	50-Mbyte (per processor)	150-Mbyte (per processor)	350-Mbyte (per processor)	1.5-Gbyte (per processor)	3.5-Gbyte (per processor)
ANY	Signature Prediction	300-900 Mbyte (per processor)	1-2.5 Gbyte (per processor)	2-6.5 Gbyte (per processor)	8.5-25 Gbyte (per processor	20-70 Gbyte (per processor)
ANY	Computational Aerodynamics	500-900 Mbyte (per processor)	1.5-2.5 Gbyte (per processor)	3.5-6.5 Gbyte (per processor)	13-25 Gbyte (per processor)	36-70 Gbyte (per processor)

PROJECT NAME: **THREAT WEAPON SYSTEMS CHARACTERIZATION**

HPC SOFTWARE REQUIREMENTS (APPLICATIONS OVERVIEW)

Major Software Application	Programming Models Used	Performance Critical Computation *(e.g., A linear solver or a sorting routine)*	Time to Solution Requirements *(S/W development time line requirements)*	Time to Solution Requirements *(Problem set-up time line requirements)*	Time to Solution Requirements *(Wall clock execution time for specified problem)*	Time to Solution Requirements *(Post run analysis time line requirements)*	Special Characteristics of Code *(e.g., Is it CPU performance, memory access, or I/O bound)*
Performance Modeling and Simulation	Various				2 days		CPU
Signature Prediction	Various				7 days		CPU
Compuational Aerodynamics	Various				18 days		CPU, Memory

PROJECT NAME: **THREAT WEAPON SYSTEMS CHARACTERIZATION**

HPC HARDWARE REQUIREMENTS (CURRENT SYSTEM BASELINE)

HPC System	Number of Computational Nodes Used by Typical Application	Number of Processors Per Node	Processor Characteristics *(e.g., Type, nominal speed)*	Memory Per Processor	I/O System Characteristics	Total Secondary Storage	Interconnect Characteristics *(Local and aggregate bandwidth, latency)*
Linux	374	2	Intel PIII/1-GHz	0.5 Gbyte	SCSI/Fast Ethernet	4-Tbyte (NAS)	Fast Ethernet
SGI O2K	7	8	MIPS R10000 250-MHz	384 Mbyte	SCSI/Fast Ethernet	4-Tbyte (NAS)	Fast Ethernet
SGI O2K	2	16	MIPS R10000 250-MHz	384 Mbyte	SCSI/Fast Ethernet	4-Tbyte (NAS)	Fast Ethernet
SGI O2K	3	32	MIPS R10000 250-MHz	368 Mbyte	SCSI/Fast Ethernet	4-Tbyte (NAS)	Fast Ethernet
Sun E420	28	4	UltraSPARC II 450-MHz	1.0 Gbyte	SCSI/Fast Ethernet	4-Tbyte (NAS)	Fast Ethernet

PROJECT NAME: **THREAT WEAPON SYSTEMS CHARACTERIZATION**

HPC SOFTWARE REQUIREMENTS (DETAIL)

Major Software Application	**Memory Access Pattern** *(e.g., Stride 1 or random)*	**Computation Profile** *(e.g., Volume of integer v. floating point computation)*	**I/O Requirements** *(e.g., Volume and rate)*	**Communication Characteristics** *(e.g., Describe message size and rate)*	**Inherent Concurrency** *(e.g., Number of independent threads in main body of calculation)*
Performance Modeling and Simulation		Predominantly Floating Point	Low		Single Threaded
Signature Prediction		Predominantly Floating Point	Low		Single Threaded
Computational Aerodynamics		Predominantly Floating Point	Low		1-8

C-11 REUSABLE LAUNCH VEHICLE (RLV) AERODATABOOK DEVELOPMENT

Project Information

HPC Project Name: Reusable Launch Vehicle (RLV) Aerodatabook Development

Agency and Organization: National Aeronautics and Space Administration

Project Leader(s): See Note [1]

Project Leader E-Mail Addres[1]s: See Note [1]

Project Leader Phone Number: See Note [1]

Project Leader Address: See Note [1]

Number of Users in Project: See Note [1]

Computational Technology Area(s) (See attached list):

Project Description

The project objective is to compute the aerodynamic and thermal load environment of the descent trajectory of a typical reusable launch vehicle configuration.

What is the real life problem?

Aerodynamic, thermal, and structural design of an RLV.

What is the National Security Impact of (solving and not solving) the problem?

There may be none.

Why is HPC needed to address the problem?

Sufficient ground test facilities to obtain the required data do not exist. This is particularly true of Mach numbers over 10.

What do you need to accomplish with HPC between 2002 and 2012?

Substantial improvements in the following areas:

- Processing speed
- Memory size
- I/O bandwidth

[1] NASA has had a number of technology programs aimed at solving certain limited elements of the RLV technology needs However, none of these programs included a full design in which a full Aerodatabook is needed. The Aerodatabook defines the aerodynamic and thermal load environment for a flight vehicle under all anticipated conditions. These data are used for detailed design, load analyses, flight simulations, and other related activities. Since none of the current programs was intended to produce a flight vehicle, there is no “NASA Project” at this time. Because of this, a full Aerodatabook ahs never been constructed. The data presented here were developed in an Agency High End Computing Requirements Study (“Strategic Assessment of NASA Requirements for High-End Computing, 1999 – 2003,” March 1998). They provide an estimate of the computing capacity required to construct a complete Aerodatabook for a single RLV configuration. Note that the data given do address Thermal Protection System (TPS) calculation requirements or any structures and materials requirements.

What do you need (H/W & S/W) from the HPC Community between 2002 and 2012 to solve your problem?

As already noted, substantially improved computational performance. In addition, to provide an iterative design capability, a dynamic, geographically-independent collaborative engineering and design capability which also includes financial estimation capability and assessment of manufacturing requirements.

Project Vision

What is the mission vision?

To provide an integrated RLV design capability.

What could you accomplish if computers were 10x more powerful?

A complete Aerodatabook for a single configuration could be computed in about 6 months instead of 5 years.

What could you accomplish if computers were 10x cheaper?

Assuming the sufficient progress was made in performance, a significant cost reduction would permit a larger, more extensive database to be computed. This would result in substantially lower risk and, perhaps, a vehicle of lower weight.

What could you accomplish if computers were 10x easier to program?

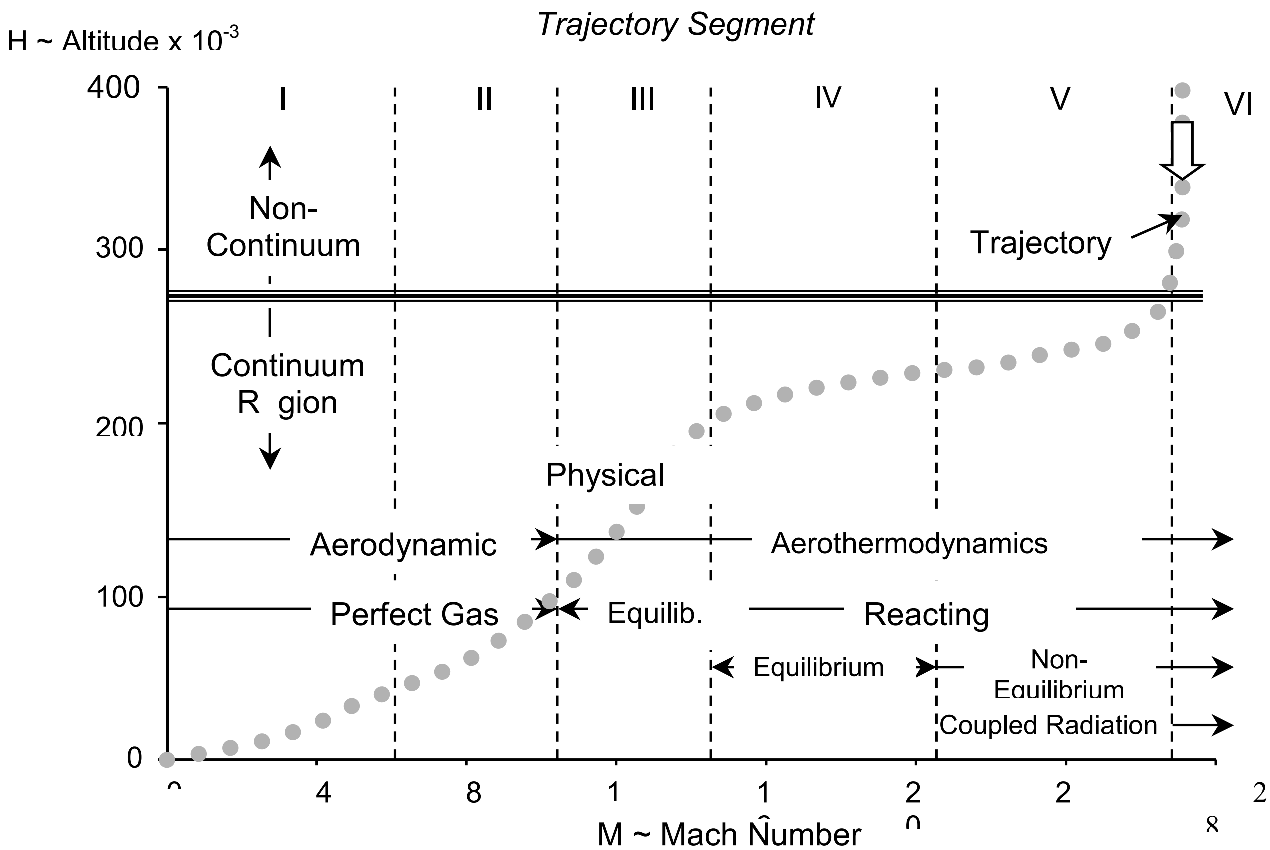
(All three of the above questions to be answered relative to progress greater than that expected from Moore's Law)

Programming is not the central issue of this requirement.

Computing the Aerodatabook for the descent trajectory of an RLV.

The diagram below presents a schematic of a typical RLV descent trajectory overlaid with the physical characteristics and types of flow and chemistry modeling required to simulate the flow in each of the six trajectory segments. To compile a complete Aerodatabook 3 trajectories are normally calculated – one for the designed trajectory and two off-design trajectories.

Typical RLV Descent Trajectory: Aerodynamic Analyses



Details of the computing required are outlined in the speadsheet on the following page.

NASA HEC Requirements Study Code R Programs RLV: Aerodatabook Requirements (no TPS/Structures/Materials)																			
Trajectory Segment		Memory, Mass Storage, & CPU Requirements																	
		per Code Run							per RLV Trajectory								per RLV Configuration		
		Memory							Code Run Multipliers										
Segment ID	Mach Range	Grid Size (Mpts)	Variables/Grid Point: Basic	Addt'l	Total	Total (MB)	Mass Stor (GB)	CPU Time (C-90 Hours)	Mach No.'s	α's	β's	Body Flap δ's	Elevon δ's	Total	Mass Stor (GB)	CPU Time (C-90 Hours)	No. of Traj.'s	Mass Stor (GB)	CPU Time (C-90 Hours)
I	0<M<6	8	8	0	8	1,280	3.2	150	5	3	3	3	3	405	1,296	60,750	3	3,888	182,250
II	6<M<10	8	8	0	8	1,280	3.2	150	1	3	3	3	3	81	259	12,150	3	778	36,450
III	10<M<14	8	8	0	8	1,280	3.2	225	1	3	1	3	3	27	86	6,075	3	259	18,225
IV	14<M<20	8	8	11	19	3,040	7.6	300	5	3	1	3	3	135	1,026	40,500	3	3,078	121,500
V	20<M<26	8	8	13	21	3,360	8.4	600	4	3	1	3	3	108	907	64,800	3	2,722	194,400
VI*	M>26	8	8	22	30	4,800	12.0	1,200	2	3	1	3	3	54	648	64,800	3	1,944	194,400

The total capacity required is then:

Total Resouces Required					
Trajectory Segment	Resource	Element: per Trajectory		per Configuration	
I - V	Mass Storage & CPU Time (Gbytes & C-90 Hrs)	3,575	184,275	10,724	552,825
	CPU Time (Equiv. von Neumann Computational Years**)		1.38		4.15
I - VI	Mass Storage & CPU Time (Gbytes & C-90 Hrs)	4,223	249,075	12,668	747,225
	CPU Time (Equiv. von Neumann Computational Years**)		1.87		5.61

** von Neumann is a 16 CPU, 1 Giga-Word Cray Research, Inc. C-90

Where the following physical modeling assumptions were made:

Trajectory Seg. ID	Physical Model Assumptions
I	Standard Aerodynamics w/Perfect Gas
II	Standard Aerodynamics w/Perfect Gas
III	Aerothermodynamics w/Equilibrium Chemistry
IV	Aerothermodynamics w/Reacting Gas/Equilibrium
V	Aerothermodynamics w/Reacting Gas/Non-Equilibrium
VI	Aerothermodynamics w/Reacting Gas/Non-Equilibrium w/Coupled Radiation